

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 19, 2004, 14:36:45 ; Search time 1815 Seconds
(without alignments)
4294.231 Million cell updates/sec

Title: US-10-051-769-2
Perfect score: 261
Sequence: 1 gatcaagtgaggatcgagg.....cactggccatcgacgtgga 261

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST.*
1: em_estba:*
2: em_estim:*
3: em_estin:*
4: em_estml:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	261	100.0	583	10 BF330450	BF330450 MR2-BN036
2	241	100.0	3069	11 BC033019	BC033019 Homo sapi
3	248	95.0	989	13 BX341552	BX341552 BX341552
4	164	62.8	944	13 BX345030	BX345030 BX345030

5	69	26.4	738	29	CC567140
6	45	17.2	437	13	BY276266
7	45	17.2	458	10	BB638573
8	45	17.2	667	13	BY734796
9	45	17.2	702	13	BU706646
10	45	17.2	930	13	BQ960960
11	45	17.2	2052	11	AK041847
12	42	16.1	723	9	AW013379
13	33	12.6	681	10	BB660050
14	33	12.6	1079	13	BX369637
15	23	8.8	464	10	BB655451
16	22	8.4	730	13	BU449181
17	22	8.4	739	13	BU463741
18	22	8.4	898	13	BU453045
19	21	8.0	290	14	CB710342
20	21	8.0	644	10	BB660976
21	21	8.0	1595	11	AK042994
22	21	8.0	2214	29	AY399665
23	21	8.0	3164	11	AK085351
24	21	8.0	3734	11	AK082438
25	20	7.7	359	13	BU282131
c 26	20	7.7	503	9	AI821366
27	20	7.7	734	13	BU404885
28	20	7.7	937	29	CNS01UUV
c 29	20	7.7	1012	13	BX343546
c 30	20	7.7	1016	9	AL537046
31	20	7.7	1033	29	CNS05MA8
c 32	19	7.3	145	12	BI262003
33	19	7.3	288	14	N43856
34	19	7.3	369	29	CG766400
c 35	19	7.3	378	29	CG985914
c 36	19	7.3	417	14	CD446645
37	19	7.3	444	12	BG314429
c 38	19	7.3	467	12	BM991757
c 39	19	7.3	473	13	BY155778
c 40	19	7.3	478	28	BH127169
c 41	19	7.3	507	9	AA587221
42	19	7.3	555	14	CD887167
c 43	19	7.3	568	12	BQ045169
c 44	19	7.3	579	9	AI295477
45	19	7.3	585	12	BJ008581

ALIGNMENTS

RESULT 1
BF330450/c
LOCUS BF330450 583 bp mRNA linear EST 22-NOV-2000
DEFINITION MR2-BN0364-280800-005-f07 BN0364 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF330450
VERSION BF330450.1 GI:11301198
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 583)
AUTHORS Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR2&t2=MR2-BN0364-280800-005-f07&t3=2000-08-28&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 10
 High quality sequence stop: 581.
 Location/Qualifiers
 1..583
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="BN0364"
 /note="Organ: breast normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

FEATURES

source

ORIGIN

Query Match 100.0%; Score 261; DB 10; Length 583;
 Best Local Similarity 100.0%; Pred. No. 4.7e-106;
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGGTGGAGTTCGAGGAGCTCTGCAGACCAGACGGCGCGCGCTGCTGGAGGG 60
 Db 482 GATCAAGGTGGAGTTCGAGGAGCTCTGCAGACCAGACGGCGCGCGCTGCTGGAGGG 423

QY 61 GCTGAGCTCGGGACGTTCTCTCGCGGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
 Db 422 GCTGAGCTCGGGACGTTCTCTCGCGGAGACGGTGCCTTCATCAAGACCATCCGGCT 363

QY 121 CGTGGCGCAGTCGTCCTCGGCACCGGGAGCCGATGCCCTGAAGGGAGCGCT 180
 Db 362 CGTGGCGCAGTCGTCCTCGGCACCGGGAGCCGATGCCCTGAAGGGAGCGCT 303

QY 181 GCCCGCCCTGCCCCGAGGAGCTGGCTTCGAGGCGAGTGGAGTACACGCGGGCTT 240
 Db 302 GCCCGCCCTGCCCCGAGGAGCTGGCTTCGAGGCGAGTGGAGTACACGCGGGCTT 243

QY 241 CCACCTGCCATCGACGTGGA 261
 Db 242 CCACCTGCCATCGACGTGGA 222

RESULT 2

BC033019
 LOCUS BC033019 3069 bp mRNA linear HTC 19-NOV-2003
 DEFINITION Homo sapiens hypothetical protein LOC118987, mRNA (cDNA clone IMAGE:4821752), containing frame-shift errors.
 ACCESSION BC033019
 VERSION BC033019.1 GI:21542573
 KEYWORDS HTC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 3069)
 REFERENCE Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Udén, T.B., Toshiyuki, S., Carninci, P., Prange, C., Rana, S.S., Loquellano, N.A., Peters, G.J.,

Abramson, R.D., Mullaby, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22388257
 MEDLINE 12477932
 PUBLISHED
 REFERENCE 2 (bases 1 to 3069)
 AUTHORS Strausberg, R.
 TITLES Direct Submission
 JOURNAL Submitted (21-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
 COMMENT Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Institute for Systems Biology
 http://www.systemsbio.org
 Contact: amadan@systemsbiology.org
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kettman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAP Plate: 33 Row: b Column: 4
 This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction
 This clone has the following problem: frame shifted.

FEATURES

source

Location/Qualifiers
 1..3069
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4821752"
 /tissue_type="Testis"
 /clone_lib="NIH MGC_97"
 /lab_host="DH10B"
 /note="Vector: pBluescript"

ORIGIN

Query Match 100.0%; Score 261; DB 11; Length 3069;
 Best Local Similarity 100.0%; Pred. No. 7e-106;
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAGACGGCGCGCTGCTGGAGGG 60
 Db 610 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAGACGGCGCGCTGCTGGAGGG 669

QY 61 GCTCAGCTCGGGACGTTCTCGGGGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
 Db 670 GCTCAGCTCGGGACGTTCTCGGGGAGACGGTGCCTTCATCAAGACCATCCGGCT 729

QY 121 CGTGGCGCCAGTTCGTCCTCGGCCACCGGGAGCCGATGGCCCTGAAGGGAGGGCT 180
 Db 730 CGTGGCGCCAGTTCGTCCTCGGCCACCGGGAGCCGATGGCCCTGAAGGGAGGGCT 789

QY 181 GCCCGCCCTGCCCCGAGGAGTGGCTTCGAGGCGGAGTGGAGTACACGCGGGCTT 240
 Db 790 GCCCGCCCTGCCCCGAGGAGTGGCTTCGAGGCGGAGTGGAGTACACGCGGGCTT 849

REFERENCE
AUTHORS 1 (bases 1 to 738)
Holt, R., Scott, J., Yang, G., Barber, S., Smailus, D., Prabhu, A., L.,
Tsai, M., Cloutier, A., Lee, D., Girn, N., Olson, T., Mayo, M.,
Butterfield, Y., Kirkpatrick, R., Liu, J., Guin, R., Chan, A., Chiu, R.,
Mathewson, C., Wye, N., Masson, A., Brown-John, M., Jones, S.,
Schein, J., Marra, M., de Jong, P., Keeler, J. W. and Kappes, S. M.
Bovine BAC End Sequences from Library CHORI-240, PLATES 399 to 478
Unpublished (2003)
Other GSSs: CH240_441O17.T7
Contact: Rob Holt
Sequencing
The British Columbia Cancer Agency Genome Science Centre
600 W. 10th Ave, Vancouver, British Columbia, Canada V5Z 4E6
Tel: 604-877-6085
Fax: 604-877-6276
Email: rholt@bcgsc.ca
Clones are derived from the bovine BAC library CHORI-240
(http://www.chori.org/bacpac/bovine240.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.choi.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/ordering/information.htm). This work
was undertaken as part of the International Bovine BAC Mapping
Consortium (IBBMC) by CSIRO Livestock Industries, Australia and the
British Columbia Genome Sciences Centre, Canada.
Plate: 441 row: O column: 17
Seq primer: SP6
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..738
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="breed: Hereford"
/db_xref="taxon:9913"
/clone="CH240_441O17"
/sex="Male"
/cell_type="Blood"
/clone_lib="CHORI-240"
/note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC
Library (Male) produced by Pieter de Jong"

ORIGIN
Query Match 26.4%; Score 69; DB 29; Length 738;
Best local Similarity 100.0%; Pred. No. 3.6e-20;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 CCCGAGGAGTGGCTTCGAGGCGGAGTGGAGTACACGGGGCTTCCACCTGGGCAT 252
Db 538 CCCGAGGAGTGGCTTCGAGGCGGAGTGGAGTACACGGGGCTTCCACCTGGGCAT 597

QY 253 CGACGTGGA 261
Db 598 CGACGTGGA 606

RESULT 6
BY276266 437 bp mRNA linear EST 11-DEC-2002
LOCUS
DEFINITION
cDNA clone K430318E15 5', mRNA sequence.
ACCESSION
BY276266
VERSION
BY276266.1 GI:26466603
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 437)
Nikaido, I., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Okazaki, Y., Furuno, M., Saito, R., Suzuki, H., Yamana, I.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Sult, C.,
Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H.,

TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
22354683
12466851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
Hayashizaki, Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Michela Fagioli and Takao K. Hensch (
Laboratory for Neuronal Circuit Development Brain Science Institute
RIKEN 2-1 Hirose, Wako-shi, Saitama 351-0198 Japan) whose
assistance we gratefully acknowledge. Please visit our web site
(http://genome.gsc.riken.go.jp) for further details.
Location/Qualifiers
1..437
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K430318E15"
/tissue_type="visual cortex"
/clone_lib="RIKEN full-length enriched, visual cortex"

FEATURES
source
ORIGIN


```

Query Match      17.2%; Score 45; DB 13; Length 437;
Best Local Similarity 100.0%; Pred. No. 1.7e-09;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGTGGAGTTCGAGGAGCTGTCGACAGCAACGCGCGG 45
    |||||
81 GATCAAGTGGAGTTCGAGGAGCTGTCGACAGCAACGCGCGG 125

RESULT 7
BBG38573
LOCUS      BBG38573      458 bp      mRNA      linear      EST 31-AUG-2001
DEFINITION BBG38573 RIKEN full-length enriched, 3 days neonate thymus Mus
            musculus cDNA clone A630041P07 5', mRNA sequence.
ACCESSION  BBG38573
VERSION     BBG38573.1 GI:15401196
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
REFERENCE  1 (bases 1 to 458)
AUTHORS    Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
            Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
            Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
            Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayaashizaki,Y.
TITLE      RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL    Unpublished (2001)
COMMENT    Contact: Yoshihide Hayaashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center (GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL: http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayaashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayaashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
            Sugahara,Y. and Hayaashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            Arakawa,K., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A.,
            Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K.,
            Arakawa,T., Ishii,Y. and Hayaashizaki,Y.
            Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.
            Func. Genomics 2 pre, 172-186 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES   e mouse tissues.
            Location/Qualifiers
            1..458
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            /clone="A630041P07"
            /tissue_type="Thymus"

```

```

/dev stage="3 days neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 3 days neonate
thymus"
/notes="Site 1: Sali; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAAGGATCCAGAGCTCTTTTTTTTTTNN 3'], cDNA was
prepared by using trihalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adaptor of
sequence [5' GAGAGAGATTCGAGTTAATTAATTCCTCCCTCCCTCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC 1."

ORIGIN
Query Match      17.2%; Score 45; DB 10; Length 458;
Best Local Similarity 100.0%; Pred. No. 1.7e-09;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGTGGAGTTCGAGGAGCTGTCGACAGCAACGCGCGG 45
    |||||
6 GATCAAGTGGAGTTCGAGGAGCTGTCGACAGCAACGCGCGG 50

RESULT 8
BY734796
LOCUS      BY734796      667 bp      mRNA      linear      EST 17-DEC-2002
DEFINITION BY734796 RIKEN full-length enriched, mammary gland RCB-0526
            YJ9-MC(A) cDNA Mus musculus cDNA clone G830026K01 5', mRNA
            sequence.
ACCESSION  BY734796
VERSION     BY734796.1 GI:27147923
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
REFERENCE  1 (bases 1 to 667)
AUTHORS    Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            NiKaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.P., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Kongaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Perteu,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramchandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Sempke,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takeraka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Alzawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayaashizaki,Y.

```

TITLE
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs

JOURNAL
Nature 420, 563-573 (2002)

MEDLINE
22354683

PUBMED
1246851

COMMENT
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P.,
Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F.,
Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,
Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,
Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N.,
Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,
Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.

FEATURES

Location/Qualifiers

```
1..667
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="G830026K01"
/tissue_type="mammary gland"
/cell_line="RCB-0526 Jyg-MC(A)"
/clone_lib="RIKEN full-length enriched, mammary gland
RCB-0526 Jyg-MC(A) cDNA"
```

ORIGIN

```
Query Match 17.2%; Score 45; DB 13; Length 667;
Best Local Similarity 100.0%; Pred. No. 1.9e-09;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 GATCAAGTGGAGTTTCGAGGAGCTGCTGCAGACCAAGACGCGCG 45

Db 556 GATCAAGTGGAGTTTCGAGGAGCTGCTGCAGACCAAGACGCGCG 600

```
RESULT 9
BU706646 702 bp mRNA linear EST 15-JUL-2003
LOCUS
DEFINITION
UI-M-FOO-cad-p-11-0-UI.r1 NIH_BMAP_F00 Mus musculus cDNA clone
IMAGE:6409978 5', mRNA sequence.
```

```
BU706646
VERSION
BU706646.1 GI:23637322
KEYWORDS
EST.
```

ORGANISM

```
Mus musculus (house mouse)
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

REFERENCE

1 (bases 1 to 702)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: pYX-5.

FEATURES

Location/Qualifiers

```
1..702
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/clone="IMAGE:6409978"
/tissue_type="whole brain"
/dev_stage="embryo 12.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_F00"
/notes="Organ: Brain; Vector: pYX-Asc; Site 1: Ecor I;  
Site 2: Not I; The library was constructed according  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a Not I site. Double strand cDNA was  
size selected according to mRNA size fraction, ligated  
with Ecor I adaptor, digested with NotI and then cloned  
directionally into pYX-Asc vector. The library tag  
sequence located between the Not I site and the polyA tail  
is TCAGAGAGCC. This library was created for the University  
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
Developing Mouse Nervous System', supported by National  
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,  
program coordinator."
```

ORIGIN

```
Query Match 17.2%; Score 45; DB 13; Length 702;
Best Local Similarity 100.0%; Pred. No. 1.9e-09;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 GATCAAGTGGAGTTTCGAGGAGCTGCTGCAGACCAAGACGCGCG 45

Db 310 GATCAAGTGGAGTTTCGAGGAGCTGCTGCAGACCAAGACGCGCG 354

RESULT 10
BQ960960

LOCUS

DEFINITION

AGENCY: 8955743 NCI CGAP_Mam2 Mus musculus cDNA clone

IMAGE:6439792 5', mRNA sequence.

BQ960960

ACCESSION

VERSION

BQ960960.1 GI:22376438

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

```

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13960 row: f column: 17
High quality sequence stop: 603.

FEATURES
    source
        1..930
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="FVB/N-3"
            /db_xref="taxon:10090"
            /clone="IMAGE:6439792"
            /tissue_type="tumor, Biopsy sample"
            /dev_stage="5 months"
            /lab_host="DH10B"
            /clone_lib="NCI_CGAP Mam2"
            /notes="Organ: mammary; Vector: pCMV-SPORT6; Site 1: Sali;
            Site 2: NotI; Cloned unidirectionally. Primer: Oligo dr.
            Library constructed by Life Technologies. Investigator
            providing samples: Gilbert Smith, NIH"

ORIGIN
Query Match      17.2%; Score 45; DB 13; Length 930;
Best Local Similarity 100.0%; Pred. No. 28-09;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCG 45
    |||||
Db 335 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCG 379

RESULT 11
AK041847
LOCUS
DEFINITION
Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
library, clone:A630041P07 product:hypothetical PDZ domain (also
known as DHR or GLGF) containing protein, full insert sequence.

ACCESSION
AK041847.1 GI:26088683
VERSION
HTC; CAP trapper.
KEYWORDS
Mus musculus (house mouse)
ORGANISM
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
MEDLINE
10149636
PUBMED

REFERENCE
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
2049374
MEDLINE
11042159
PUBMED

REFERENCE
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitzunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, K., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multipipillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
MEDLINE
11076861
PUBMED

AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitzunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, K., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

TITLE
hypothesized PDZ domain (also known as DHR or GLGF)
containing protein (InterPro|IPR001478, evidence:
InterPro)

ORIGIN
Query Match      17.2%; Score 45; DB 11; Length 2052;
Best Local Similarity 100.0%; Pred. No. 2.5e-09;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCG 45
    |||||
Db 10 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCG 54

RESULT 12
AW013379

```

```

4 The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

5 The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

6 (bases 1 to 2052)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.

Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome
cDNA library Project of Genome Exploration Research Group in Riken
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
Trust/MRC building Addenbrookes Hospital Cambridge) whose
assistance we gratefully acknowledge.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.
Location/Qualifiers
1..2052
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:A630041P07"
/db_xref="WGI:2406859"
/db_xref="taxon:10090"
/clone="A630041P07"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="3 days neonate"

misc_feature
1..2052
/notes="hypothesized PDZ domain (also known as DHR or GLGF)
containing protein (InterPro|IPR001478, evidence:
InterPro)"

ORIGIN
Query Match      17.2%; Score 45; DB 11; Length 2052;
Best Local Similarity 100.0%; Pred. No. 2.5e-09;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCG 45
    |||||
Db 10 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCG 54

RESULT 12
AW013379

```

LOCUS AW013379 723 bp mRNA linear **EST 10-SEP-1999**
 DEFINITION sp042ks winter flounder spleen Pseudopleuronectes americanus cDNA clone sp042ks 5' similar to C53B4.4 [Caenorhabditis elegans], mRNA sequence.

ACCESSION AW013379
 VERSION AW013379.1 GI:5862157
 KEYWORDS EST.
 ORGANISM Pseudopleuronectes americanus (winter flounder)

REFERENCE 1 (bases 1 to 723)
 AUTHORS Douglas, S.E., Gallant, J.W., Bullerwell, C.E., Wolff, C., Munholland, J., and Reith, M.E.
 TITLE Winter flounder expressed sequence tags: Establishment of an EST database and identification of novel fish genes
 JOURNAL Marine Biotechnology (1999) In press
 COMMENT Contact: Reith M
 Marine Biology
 NRC Institute for Marine Biosciences
 1411 Oxford St., Halifax, Nova Scotia, B3H 3Z1, Canada
 Tel: (902) 426-8276
 Fax: (902) 426-9413
 Email: michael.reith@nrc.ca
 Seq primer: M13 Forward.

FEATURES
 Location/Qualifiers
 1..723
 /organism="Pseudopleuronectes americanus"
 /mol_type="mRNA"
 /db_xref="taxon:8265"
 /clone="sp042ks"
 /sex="female"
 /dev_stage="adult"
 /clone_lib="Winter flounder spleen"
 /note="Organ: spleen"

ORIGIN
 Query Match 16.1%; Score 42; DB 9; Length 723;
 Best Local Similarity 100.0%; Pred. No. 4.2e-08;
 Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGC 42
 |||||
 DB 482 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGC 523
 |||||

RESULT 13
 BB866050
 LOCUS BB866050 RIKEN full-length enriched, CRL-1751 WEHI 164 cDNA Mus
 DEFINITION musculus cDNA clone G431003009 5', mRNA sequence.

ACCESSION BB866050
 VERSION BB866050.1 GI:17112260
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 681)
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Akimura, T., Arawaka, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akaira, S., Tanaka, T., Tomaru, A., Toya, T., Watahiki, A., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)
 Unpublished (2001)

COMMENT

Contact: Yoshinide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

FEATURES

Location/Qualifiers

1..681
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="BALB/c"
 /db_xref="taxon:10090"
 /clone="G431003009"
 /cell_line="CRL-1751 WEHI 164"
 /clone_lib="RIKEN full-length enriched, CRL-1751 WEHI 164 cDNA"

ORIGIN

Query Match 12.6%; Score 33; DB 10; Length 681;
 Best Local Similarity 100.0%; Pred. No. 0.00044;
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GGTGAGTTCGAGGAGCTGCTGCAGACCAAGAC 39
 |||||
 DB 565 GGTGAGTTCGAGGAGCTGCTGCAGACCAAGAC 597
 |||||

RESULT 14
 BX369637

LOCUS BX369637 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
 DEFINITION cDNA clone CS0DK002YA12 5-PRIME, mRNA sequence.

ACCESSION BX369637
 VERSION BX369637.1 GI:30453826
 KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1079)

AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

TITLE Full-length cDNA libraries and normalization

JOURNAL Unpublished (2001)

COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 5483.r For

more information about this cluster, see

source

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 14:47:40 ; Search time 56 Seconds
(without alignments)
2586.468 Million cell updates/sec

Title: US-10-051-769-2
Perfect score: 261
Sequence: 1 gatcaagggtgagttcgagg.....cacctggccatcacgtgga 261

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	261	100.0	4303	4	US-09-976-594-899
2	21	8.0	1425	2	US-08-356-060A-6
3	21	8.0	1425	3	US-08-460-900C-6
4	21	8.0	1425	3	US-08-674-509B-6
5	21	8.0	1425	3	US-08-954-698-6
6	21	8.0	1425	4	US-08-957-874-6
7	21	8.0	1425	4	US-09-325-256-10
8	21	8.0	1425	4	US-09-639-695-6
9	21	8.0	1425	4	US-09-448-188-6
10	21	8.0	1425	4	US-08-954-128-6
11	21	8.0	1425	4	US-09-704-917-6
12	21	8.0	1425	4	US-08-954-740-6
13	21	8.0	1425	4	US-09-151-999-6
14	21	8.0	1425	4	US-09-736-476-6
15	21	8.0	1576	1	US-08-748-591-5
16	21	8.0	1576	1	US-08-748-591-10
17	21	8.0	2274	4	US-09-772-647-3
18	18	6.9	96	3	US-08-484-322-24
19	18	6.9	402	2	US-08-193-078B-20
20	18	6.9	3225	1	US-08-306-691B-45
21	18	6.9	3225	5	PCT-US93-06251-91
22	17	6.5	291	4	US-09-252-991A-10114
23	17	6.5	390	4	US-09-252-991A-4800
24	17	6.5	552	4	US-09-252-991A-4575
25	17	6.5	594	4	US-09-252-991A-9966
26	17	6.5	648	4	US-09-252-991A-4749
27	17	6.5	681	4	US-09-252-991A-4694

c	28	17	6.5	1227	4	US-09-252-991A-4732	Sequence 4732, Ap
c	29	17	6.5	1245	4	US-09-252-991A-4359	Sequence 4359, Ap
c	30	17	6.5	1305	4	US-09-252-991A-4477	Sequence 4477, Ap
c	31	17	6.5	1350	4	US-09-252-991A-10184	Sequence 10184, A
c	32	17	6.5	1374	4	US-09-252-991A-4384	Sequence 4384, Ap
c	33	17	6.5	1678	6	5223391-1	Patent No. 5223391
c	34	17	6.5	3639	4	US-09-252-991A-10045	Sequence 10045, A
c	35	17	6.5	3708	2	US-08-680-326-29	Sequence 29, Appl
c	36	17	6.5	3708	4	US-09-904-065-5	Sequence 5, Appl
c	37	17	6.5	3708	4	US-09-904-065-7	Sequence 7, Appl
c	38	17	6.5	3708	4	US-09-904-065-9	Sequence 9, Appl
c	39	17	6.5	3717	4	US-09-252-991A-9792	Sequence 9792, Ap
c	40	17	6.5	3717	4	US-09-904-065-1	Sequence 1, Appl
c	41	17	6.5	3723	4	US-09-904-065-3	Sequence 3, Appl
c	42	17	6.5	4057	3	US-08-894-997-49	Sequence 49, Appl
c	43	17	6.5	4941	4	US-09-252-991A-5741	Sequence 5741, Ap
c	44	17	6.5	23673	4	US-09-773-816-1	Sequence 1, Appl
c	45	17	6.5	4403765	3	US-09-103-840A-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-976-594-899
; Sequence 899, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 899
; LENGTH: 4303
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. 6673549 225420.1
US-09-976-594-899

Query Match	100.0%	Score 261;	DB 4;	Length 4303;
Best Local Similarity	100.0%	Pred. No. 2.5e-110;		
Matches 261;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	GATCAAGGTGAGTTCGAGGAGCTGTGCAGACCAAGACGCGCGCGCCCTGCTGGAGGG	60	
Db	79	GATCAAGGTGAGTTCGAGGAGCTGTGCAGACCAAGACGCGCGCGCCCTGCTGGAGGG	138	
Qy	61	GCTGAGCTCGGAGACCTGTTCTTGGCGAGACCGTTCCTTCATCAGACCATCCGCT	120	
Db	139	GCTGAGCTCGGAGACCTGTTCTTGGCGAGACCGTTCCTTCATCAGACCATCCGCT	198	
Qy	121	CGTGGCGGACGTCTGTCCTCGGCACCGGGAGCCCGATGGCCCTGAAGGGGAGCGCT	180	
Db	199	CGTGGCGGACGTCTGTCCTCGGCACCGGGAGCCCGATGGCCCTGAAGGGAGCGCT	258	
Qy	181	GCCCGCGCTGCCCGAGGAGCTGGCTTCGAGCGGAGGTGAGTGAACCGGGGCTT	240	
Db	259	GCCCGCGCTGCCCGAGGAGCTGGCTTCGAGCGGAGGTGAGTGAACCGGGGCTT	318	
Qy	241	CCACTGGCCATCGACGTGGA	261	
Db	319	CCACTGGCCATCGACGTGGA	339	
RESULT 2				

US-08-356-060A-6
; Sequence 6, Application US/08356060A
; Patent No. 5844079
; GENERAL INFORMATION:
; APPLICANT: Ingham, Phillip W.
; APPLICANT: McMahon, Andrew P.
; APPLICANT: Tabin, Clifford J.
; TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing
; TITLE OF INVENTION: Proteins and Uses Related Thereto
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/356,060A
; FILING DATE: 14-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/176,427
; FILING DATE: 30-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: HMI-006CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1425 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1425
US-08-356-060A-6

Query Match 8.0%; Score 21; DB 2; Length 1425;
Best Local Similarity 100.0%; Pred. No. 0.87;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 CTCGGCCACCGGGAGCCCGA 159
Db 837 CTCGGCCACCGGGAGCCCGA 857

RESULT 3

US-08-460-900C-6
; Sequence 6, Application US/08460900C
; Patent No. 6165747
; GENERAL INFORMATION:
; APPLICANT: Ingham, Phillip W.
; APPLICANT: McMahon, Andrew P.
; APPLICANT: Tabin, Clifford J.
; APPLICANT: Bumcrot, David A.
; APPLICANT: Marti-Gorostiza, Elisa
; TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing
; TITLE OF INVENTION: Proteins and Uses Related Thereto
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square

; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,900C
; FILING DATE: 5-JUNE-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/435,093
; FILING DATE: 4-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/356,060
; FILING DATE: 14-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/176,427
; FILING DATE: 30-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: HMY-006.05
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 832-1000
; TELEFAX: (617) 832-7000
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1425 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1425
US-08-460-900C-6

Query Match 8.0%; Score 21; DB 3; Length 1425;
Best Local Similarity 100.0%; Pred. No. 0.87;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 CTCGGCCACCGGGAGCCCGA 159
Db 837 CTCGGCCACCGGGAGCCCGA 857

RESULT 4

US-08-674-509B-6
; Sequence 6, Application US/08674509B
; Patent No. 6261786
; GENERAL INFORMATION:
; APPLICANT: Ingham, Phillip W.
; APPLICANT: McMahon, Andrew P.
; APPLICANT: Tabin, Clifford J.
; APPLICANT: Mariago, Valeria
; TITLE OF INVENTION: SCREENING ASSAYS FOR HEDGEGOG AGONISTS
; TITLE OF INVENTION: AND ANTAGONISTS
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30

;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/674,509B
;/ FILING DATE: 02-JUL-1996
;/ CLASSIFICATION: 435
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US 08/460,900
;/ FILING DATE: 05-JUN-1995
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Vincent, Matthew P.
;/ REGISTRATION NUMBER: 36,709
;/ REFERENCE/DOCKET NUMBER: HMV-006.06
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: 617-832-1000
;/ TELEFAX: 617-832-7000
;/ INFORMATION FOR SEQ ID NO: 6:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 1425 base pairs
;/ TYPE: nucleic acid
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: cDNA
;/ FEATURE:
;/ NAME/KEY: CDS
;/ LOCATION: 1..1425
;/ US-08-674-509B-6

Query Match 8.0%; Score 21; DB 3; Length 1425;
Best Local Similarity 100.0%; Pred. No. 0.87;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 CTCGGCCACCGGGAGCCCGA 159
|||
Db 837 CTCGGCCACCGGGAGCCCGA 857

RESULT 5
US-08-954-698-6
; Sequence 6, Application US/08954698
; Patent No. 6271363
; GENERAL INFORMATION:
; APPLICANT: Ingham, Phillip W.
; APPLICANT: McMahon, Andrew P.
; APPLICANT: Tabin, Clifford J.
; TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing
; TITLE OF INVENTION: Proteins and Uses Related Thereto
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/954,698
; FILING DATE: 20-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/462,386
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/435,093
; FILING DATE: 04-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/356,060
; FILING DATE: 14-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: HMV-006.06
; FILING DATE: 14-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/176,427
; FILING DATE: 30-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: HMV-006.06
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 832-1000
; TELEFAX: (617) 832-7000

;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Vincent, Matthew P.
;/ REGISTRATION NUMBER: 36,709
;/ REFERENCE/DOCKET NUMBER: HMV-006.10
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: 617-832-1000
;/ TELEFAX: 617-832-7000
;/ INFORMATION FOR SEQ ID NO: 6:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 1425 base pairs
;/ TYPE: nucleic acid
;/ STRANDEDNESS: both
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: cDNA
;/ FEATURE:
;/ NAME/KEY: CDS
;/ LOCATION: 1..1425
;/ US-08-954-698-6

Query Match 8.0%; Score 21; DB 3; Length 1425;
Best Local Similarity 100.0%; Pred. No. 0.87;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 CTCGGCCACCGGGAGCCCGA 159
|||
Db 837 CTCGGCCACCGGGAGCCCGA 857

RESULT 6
US-08-957-874-6
; Sequence 6, Application US/08957874
; Patent No. 6384192
; GENERAL INFORMATION:
; APPLICANT: Ingham, Phillip W.
; APPLICANT: McMahon, Andrew P.
; APPLICANT: Tabin, Clifford J.
; TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing
; TITLE OF INVENTION: Proteins and Uses Related Thereto
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,874
; FILING DATE: 20-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/462,386
; FILING DATE: 5-JUNE-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/435,093
; FILING DATE: 4-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/356,060
; FILING DATE: 14-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/176,427
; FILING DATE: 30-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: HMV-006.09
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 832-1000
; TELEFAX: (617) 832-7000

; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1425 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1..1425
 ; US-08-874-6

Query Match 8.0%; Score 21; DB 4; Length 1425;
 Best Local Similarity 100.0%; Pred. No. 0.87;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 CTCGCCACCGGGAGCCCGA 159
 |||||
 DB 837 CTCGCCACCGGGAGCCCGA 857

RESULT 7
 US-09-325-256-10
 ; Sequence 10, Application US/09325256
 ; Patent No. 6444793
 ; GENERAL INFORMATION:
 ; APPLICANT: PEPINSKY, R. BLAKE
 ; APPLICANT: BAKER, DARREN P.
 ; APPLICANT: WEN, DINGYI
 ; APPLICANT: WILLIAMS, KEVIN P.
 ; APPLICANT: GARGER, EILEEN A.
 ; APPLICANT: TAYLOR, FREDERICK R.
 ; APPLICANT: GALDES, ALPHONSE
 ; APPLICANT: PORTER, JEFFREY
 ; TITLE OF INVENTION: HYDROPHOBICALLY-MODIFIED PROTEIN COMPOSITIONS AND
 ; FILE REFERENCE: BIV-067.01
 ; CURRENT FILING DATE: 1999-06-03
 ; PRIOR APPLICATION NUMBER: US/09/325,256
 ; PRIOR FILING DATE: 1998-09-10
 ; PRIOR APPLICATION NUMBER: 60/099,800
 ; PRIOR FILING DATE: 1998-03-20
 ; PRIOR APPLICATION NUMBER: 60/078,935
 ; PRIOR FILING DATE: 1998-06-17
 ; PRIOR APPLICATION NUMBER: 60/089,685
 ; PRIOR FILING DATE: 1997-12-03
 ; PRIOR APPLICATION NUMBER: 60/067,423
 ; PRIOR FILING DATE: 1998-12-03
 ; PRIOR APPLICATION NUMBER: PCT/US98/25676
 ; NUMBER OF SEQ ID NOS: 31
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 10
 ; LENGTH: 1425
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1425)
 ; US-09-325-256-10

Query Match 8.0%; Score 21; DB 4; Length 1425;
 Best Local Similarity 100.0%; Pred. No. 0.87;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 CTCGCCACCGGGAGCCCGA 159
 |||||
 DB 837 CTCGCCACCGGGAGCCCGA 857

RESULT 8
 US-09-639-695-6
 ; Sequence 6, Application US/09639695

; Patent No. 6576237
 ; GENERAL INFORMATION:
 ; APPLICANT: Ingham, Phillip W.
 ; McMahon, Andrew P.
 ; Tabin, Clifford J.
 ; Bumcrot, David A. Elisa
 ; Marti-Gorostiza, Elisa
 ; TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing
 ; Proteins and Uses Related Thereto
 ; NUMBER OF SEQUENCES: 62
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
 ; STREET: One Post Office Square
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/639,695
 ; FILING DATE: 16-Aug-2000
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/435,093
 ; FILING DATE: 4-MAY-1995
 ; APPLICATION NUMBER: US 08/356,060
 ; FILING DATE: 14-DEC-1994
 ; APPLICATION NUMBER: US 08/176,427
 ; FILING DATE: 30-DEC-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Vincent, Matthew P.
 ; REGISTRATION NUMBER: 36,709
 ; REFERENCE/DOCKET NUMBER: HMV-006.05
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 832-1000
 ; TELEFAX: (617) 832-7000
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1425 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1..1425
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
 ; US-09-639-695-6
 ; Query Match 8.0%; Score 21; DB 4; Length 1425;
 ; Best Local Similarity 100.0%; Pred. No. 0.87;
 ; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 ; QY 139 CTCGCCACCGGGAGCCCGA 159
 ; |||||
 ; DB 837 CTCGCCACCGGGAGCCCGA 857
 ; RESULT 9
 ; US-09-448-188-6
 ; Sequence 6, Application US/09448188
 ; Patent No. 6607913
 ; GENERAL INFORMATION:
 ; APPLICANT: Ingham, Phillip W.
 ; McMahon, Andrew P.
 ; Tabin, Clifford J.
 ; TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing
 ; Proteins and Uses Related Thereto
 ; NUMBER OF SEQUENCES: 48
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/448,188
FILING DATE: 23-NOV-1999
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/462,386
FILING DATE: 05-JUN-1995
APPLICATION NUMBER: US 08/435,093
FILING DATE: 04-MAY-1995
APPLICATION NUMBER: US 08/356,060
FILING DATE: 14-DEC-1994
APPLICATION NUMBER: US 08/176,427
FILING DATE: 30-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: HMV-006.12
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-1000
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1425 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1425
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-448-188-6

Query Match 8.0%; Score 21; DB 4; Length 1425;
Best Local Similarity 100.0%; Pred. No. 0.87;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 CTCGGCCACCGGGAGCCCGA 159
|||||
DB 837 CTCGGCCACCGGGAGCCCGA 857

RESULT 10
US-09-448-188-6
Sequence 6, Application US/08954128
Patent No. 6610656
GENERAL INFORMATION:
APPLICANT: Ingham, Phillip W.
APPLICANT: McMahon, Andrew P.
APPLICANT: Tabin, Clifford J.
TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/954,128
FILING DATE: 20-OCT-1997
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/462,386
FILING DATE: 05-JUN-1995
APPLICATION NUMBER: US 08/435,093
FILING DATE: 04-MAY-1995
APPLICATION NUMBER: US 08/356,060
FILING DATE: 14-DEC-1994
APPLICATION NUMBER: US 08/176,427
FILING DATE: 30-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: HMV-006.12
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1425 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1425
US-08-954-128-6

Query Match 8.0%; Score 21; DB 4; Length 1425;
Best Local Similarity 100.0%; Pred. No. 0.87;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 CTCGGCCACCGGGAGCCCGA 159
|||||
DB 837 CTCGGCCACCGGGAGCCCGA 857

RESULT 11
US-09-704-917-6
Sequence 6, Application US/09704917
Patent No. 6616926
GENERAL INFORMATION:
APPLICANT: Biogen, Inc.
APPLICANT: Burkly, Linda
APPLICANT: Wang, Li Chun
TITLE OF INVENTION: METHODS OF MODULATING LIPID METABOLISM AND STORAGE
FILE REFERENCE: A069PCT
CURRENT APPLICATION NUMBER: US/09/704,917
CURRENT FILING DATE: 2000-11-02
PRIOR APPLICATION NUMBER: 60/122,640
PRIOR FILING DATE: 1999-03-03
PRIOR APPLICATION NUMBER: 60/124,446
PRIOR FILING DATE: 1999-03-15
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 1425
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1425)
OTHER INFORMATION: "nnn" encoding "Xaa" at position 1387-1389 may be a, t, c,
OTHER INFORMATION: g, other or unknown
US-09-704-917-6

Query Match 8.0%; Score 21; DB 4; Length 1425;
 Best Local Similarity 100.0%; Pred. No. 0.87;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 CTCGGCCACCGGGAGCCCGA 159
 |||||
 Db 837 CTCGGCCACCGGGAGCCCGA 857

RESULT 12
 US-08-954-740-6
 ; Sequence 6, Application US/08954740
 ; Patent No. 6630148
 ; GENERAL INFORMATION:
 ; APPLICANT: Ingham, Phillip W.
 ; APPLICANT: McMahon, Andrew P.
 ; APPLICANT: Tabin, Clifford J.
 ; TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing
 ; TITLE OF INVENTION: Proteins and Uses Related Thereto
 ; NUMBER OF SEQUENCES: 48
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
 ; STREET: One Post Office Square
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02109-2170
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/954,740
 ; FILING DATE: 20-OCT-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/462,386
 ; FILING DATE: 05-JUN-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/435,093
 ; FILING DATE: 04-MAY-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/356,060
 ; FILING DATE: 14-DEC-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/176,427
 ; FILING DATE: 30-DEC-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Vincent, Matthew P.
 ; REGISTRATION NUMBER: 36,709
 ; REFERENCE/DOCKET NUMBER: HMV-006.08
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-832-1000
 ; TELEFAX: 617-832-7000
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1425 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: both
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: CDNA
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1..1425
 ; US-08-954-740-6

Query Match 8.0%; Score 21; DB 4; Length 1425;
 Best Local Similarity 100.0%; Pred. No. 0.87;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 CTCGGCCACCGGGAGCCCGA 159
 |||||

Db 837 CTCGGCCACCGGGAGCCCGA 857

RESULT 13
 US-09-151-999-6
 ; Sequence 6, Application US/09151999
 ; Patent No. 6639051
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Elizabeth
 ; TITLE OF INVENTION: REGULATION OF EPITHELIAL TISSUE BY HEDGEHOG-LIKE
 ; TITLE OF INVENTION: POLYPEPTIDES, AND FORMULATIONS AND USES RELATED THERETO
 ; FILE REFERENCE: ONV-031.02
 ; CURRENT APPLICATION NUMBER: US/09/151,999
 ; CURRENT FILING DATE: 1998-08-11
 ; EARLIER APPLICATION NUMBER: 08/955,552
 ; EARLIER FILING DATE: 1997-10-20
 ; NUMBER OF SEQ ID NOS: 28
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 6
 ; LENGTH: 1425
 ; TYPE: DNA
 ; ORGANISM: Homo sapien Shh
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1425)
 ; FEATURE:
 ; OTHER INFORMATION: "nnn" encoding "Xaa" at position 1387-1389 may be a, t, c,
 ; OTHER INFORMATION: g, other or unknown
 ; US-09-151-999-6

Query Match 8.0%; Score 21; DB 4; Length 1425;
 Best Local Similarity 100.0%; Pred. No. 0.87;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 CTCGGCCACCGGGAGCCCGA 159
 |||||
 Db 837 CTCGGCCACCGGGAGCCCGA 857

RESULT 14
 US-09-736-476-6
 ; Sequence 6, Application US/09736476
 ; Patent No. 6664075
 ; GENERAL INFORMATION:
 ; APPLICANT: Ingham, Phillip W.
 ; McMahon, Andrew P.
 ; Tabin, Clifford J.
 ; Bumerot, David A.
 ; Marti-Gorostiza, Elisa
 ; TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing
 ; TITLE OF INVENTION: Proteins and Uses Related Thereto
 ; NUMBER OF SEQUENCES: 47
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD
 ; STREET: 60 State Street
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: ASCII(text)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/736,476
 ; FILING DATE: 13-Dec-2000
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/435,093
 ; FILING DATE: 4-MAY-1995
 ; APPLICATION NUMBER: US 08/356,060
 ; FILING DATE: 14-DEC-1994

```

; APPLICATION NUMBER: US 08/176,427
; FILING DATE: 30-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: HMI-006CF4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1425 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1425
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-736-476-6

Query Match      8.0%; Score 21; DB 4; Length 1425;
Best Local Similarity 100.0%; Pred. No. 0.87; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      139 CTGGGCCACCGGGGAGCCCGA 159
      |||||||||||||||||||
Db      837 CTGGGCCACCGGGGAGCCCGA 857

RESULT 15
US-08-748-591-5
; Sequence 5, Application US/08748591
; Patent No. 5759811
; GENERAL INFORMATION:
; APPLICANT: Epstein, Ervin
; APPLICANT: Hu, Zhilan
; APPLICANT: Bonifas, Jeanette
; TITLE OF INVENTION: Mutant Human Hedgehog Gene
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish and Richardson
; STREET: 2200 Sand Hill Road
; CITY: Menlo Park
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/748,591
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: 06510/067001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 322-5070
; TELEFAX: (415) 854-0875
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-748-591-5
```

```

Query Match      8.0%; Score 21; DB 1; Length 1576;
Best Local Similarity 100.0%; Pred. No. 0.86; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      139 CTGGGCCACCGGGGAGCCCGA 159
      |||||||||||||||||||
Db      988 CTGGGCCACCGGGGAGCCCGA 1008

Search completed: July 18, 2004, 15:54:55
Job time : 58 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 12:32:20 ; Search time 88.141 Seconds
(without alignments)
1643.301 Million cell updates/sec

Title: US-10-051-769-2
Perfect score: 261
Sequence: 1 gatcaagtgaggtcgagg.....cactggccatgcagctgga 261

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/prodata/2/ina/5A COMB.seq: *
2: /cgn2_6/prodata/2/ina/5B COMB.seq: *
3: /cgn2_6/prodata/2/ina/6A COMB.seq: *
4: /cgn2_6/prodata/2/ina/6B COMB.seq: *
5: /cgn2_6/prodata/2/ina/PCUS COMB.seq: *
6: /cgn2_6/prodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	261	100.0	4303	4	US-09-976-594-899 Sequence 899, App
2	51.6	19.8	6436	4	US-09-600-099-1 Sequence 1, Appli
3	44.4	17.0	450	4	US-09-252-991A-8011 Sequence 8011, Ap
4	44.4	17.0	1467	4	US-09-252-991A-8188 Sequence 8188, Ap
5	44.4	17.0	2823	4	US-09-252-991A-8068 Sequence 8068, Ap
6	44.4	17.0	2949	4	US-09-252-991A-8068 Sequence 8249, Ap
7	44.4	17.0	23673	4	US-09-773-816-1 Sequence 1, Appli
8	43.6	16.7	4849	4	US-09-620-312D-39 Sequence 39, Appl
9	43.4	16.6	4403765	3	US-09-103-840A-2 Sequence 2, Appli
10	43.4	16.6	4411529	3	US-09-103-840A-1 Sequence 1, Appli
11	43.2	16.6	2825	4	US-09-196-390-5 Sequence 5, Appli
12	42.4	16.2	77536	4	US-09-410-551B-1 Sequence 1, Appli
13	42.2	16.2	450	4	US-09-252-991A-10221 Sequence 10221, A
14	42.2	16.2	561	4	US-09-252-991A-10438 Sequence 10438, A
15	42.2	16.2	705	4	US-09-252-991A-10340 Sequence 10340, A
16	42.2	16.2	1053	4	US-09-252-991A-13541 Sequence 13541, A
17	42.2	16.2	1167	4	US-09-252-991A-10810 Sequence 10810, A
18	42.2	16.2	1157	4	US-09-252-991A-13907 Sequence 13907, A
19	42.2	16.2	1722	4	US-09-252-991A-13907 Sequence 13800, A
20	41.8	16.0	3937	4	US-10-164-595-7 Sequence 7, Appli
21	41.6	15.9	1910	3	US-09-593-711A-3 Sequence 3, Appli
22	41.6	15.9	1914	1	US-07-601-094-1 Sequence 1, Appli
23	41.6	15.9	1914	1	US-08-012-735-1 Sequence 1, Appli
24	41.4	15.9	1194	4	US-09-252-991A-11352 Sequence 11352, A
25	41.4	15.9	1500	4	US-09-252-991A-11312 Sequence 11312, A
26	41.4	15.9	2618	4	US-09-857-556A-25 Sequence 25, Appl
27	41	15.7	1398	4	US-09-252-991A-14375 Sequence 14375, A

C	28	41	15.7	1416	4	US-09-252-991A-14328	Sequence 14328, A
	29	40.8	15.6	942	2	US-08-446-806-2	Sequence 2, Appli
	30	40.8	15.6	942	3	US-09-385-028-17	Sequence 17, Appl
	31	40.8	15.6	942	4	US-09-726-614-17	Sequence 17, Appl
	32	40.8	15.6	942	4	US-09-385-040-17	Sequence 17, Appl
C	33	40.8	15.6	1218	4	US-09-252-991A-7267	Sequence 7267, Ap
	34	40.8	15.6	11604	3	US-09-385-028-13	Sequence 13, Appl
	35	40.8	15.6	11604	4	US-09-726-614-13	Sequence 13, Appl
	36	40.8	15.6	11604	4	US-09-385-040-13	Sequence 13, Appl
	37	40.8	15.6	15079	3	US-09-385-028-1	Sequence 1, Appli
	38	40.8	15.6	15079	4	US-09-726-614-1	Sequence 1, Appli
	39	40.8	15.6	15120	4	US-09-385-040-1	Sequence 1, Appli
	40	40.4	15.5	513	4	US-09-252-991A-9735	Sequence 9735, Ap
C	41	40.4	15.5	858	4	US-09-252-991A-9510	Sequence 9510, Ap
	42	40.4	15.5	918	4	US-09-252-991A-9653	Sequence 9653, Ap
	43	40.4	15.5	1062	4	US-09-252-991A-9694	Sequence 9694, Ap
	44	40.4	15.5	1953	4	US-09-252-991A-5663	Sequence 5663, Ap
C	45	40.4	15.5	2721	4	US-09-252-991A-5561	Sequence 5561, Ap

ALIGNMENTS

RESULT 1
US-09-976-594-899
; Sequence 899, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 899
; LENGTH: 4303
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. 6673549 225420.1
US-09-976-594-899

Query Match	100.0%	Score 261;	DB 4;	Length 4303;
Best Local Similarity	100.0%;	Pred. No. 5.5e-48;	Mismatches 0;	Indels 0;
Matches 261;	Conservative 0;	Indels 0;	Gaps 0;	
QY	1	GATCAAGGTGGAGTTCGAGGAGCTGTCAGACCAAGACGGCGGGCGCTGCTGGAGGG	60	
Db	79	GATCAAGGTGGAGTTCGAGGAGCTGTCAGACCAAGACGGCGGGCGCTGCTGGAGGG	138	
QY	61	GCTGAGCTCGGGAGCGTTCCTGGGCGAGACGGTGCCTTCATCAGACCATCCGGCT	120	
Db	139	GCTGAGCTCGGGAGCGTTCCTGGGCGAGACGGTGCCTTCATCAGACCATCCGGCT	198	
QY	121	CCTGGCGCCAGTCGTGCGCCCTCGGCACCGGGAGCCCGATGCGCCCTGAAGGGAGCGCT	180	
Db	199	CCTGGCGCCAGTCGTGCGCCCTCGGCACCGGGAGCCCGATGCGCCCTGAAGGGAGCGCT	258	
QY	181	GCCCGCCGCTCCCGGAGAGCTGGCCCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT	240	
Db	259	GCCCGCCGCTCCCGGAGAGCTGGCCCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT	318	
QY	241	CCACCTGGCCATCGAGCTGGA	261	
Db	319	CCACCTGGCCATCGAGCTGGA	339	
RESULT 2				


```

RESULT 8
US-09-620-312D-39
; Sequence 39, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenchua

```


QY 218 GAGGTGGAG 226
Db 601015 ATGGGCGG 601023

RESULT 11
US-09-196-390-5
; Sequence 5, Application US/09196390
; Patent No. 6307125
; GENERAL INFORMATION:
; APPLICANT: Block, Martina
; APPLICANT: Lorz, Horst
; APPLICANT: Lutticke, Stephanie
; APPLICANT: Walter, Lennart
; APPLICANT: Froberg, Claus
; APPLICANT: Kossmann, Jens
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING ENZYMES
; TITLE OF INVENTION: FROM WHEAT WHICH ARE INVOLVED IN STARCH
; TITLE OF INVENTION: SYNTHESIS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James F. Haley, Jr., c/o Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,390
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 21 588.9
; FILING DATE: 29-MAY-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 36 917.7
; FILING DATE: 11-SEP-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP97/02793
; FILING DATE: 28-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: AGREVO-9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2825 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Triticum aestivum L.
; STRAIN: cv. Florida
; TISSUE TYPE: ca. 21 d Caryopses
; IMMEDIATE SOURCE:
; LIBRARY: cDNA library in pBluescript sk (-)
; CLONE: pTASS1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 162..2559
US-09-196-390-5

Query Match 16.6%; Score 43.2; DB 4; Length 2825;
Best Local Similarity 51.6%; Pred. No. 0.39;
Matches 99; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 10 GGAGTTTCAGGAGCTGCTGCAGACCAAGACGGCGCGGCTGCTGTGAGGGGCTGAGCCT 69
Db 2141 GCACCTTCAGCGGAGACCAACAGAGGTGCGGGGTGGGTTCCTCCGTGCGCCT 2200

QY 70 GCGGACGTCTTCTGCGGAGACGGTGCCTTTCATCAAGACCATCCGGCTCGTGGCGCC 129
Db 2201 GCGGACCGGATCACGGCGGGGCGGACGCTCTCATCCCTCCCGTTCGAGCGCGTG 2260

QY 130 AGTCGTGCTCCGCGACCGGGAGCCCGATGCGCTGAAGGGAGGCGCTGCCCGCGC 189
Db 2261 CGGCTGAACCACTCTACGCCATGGCTACGGCACCGTCCCGTCGTGCACGCGTCGG 2320

QY 190 CTGCCCGGAGGA 201
Db 2321 CGGCTCAGGGA 2332

RESULT 12
US-09-410-551B-1/c
; Sequence 1, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; TITLE OF INVENTION: CONSTRUCTS THEREFOR
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 77536
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52275)...(71465)
US-09-410-551B-1

Query Match 16.2%; Score 42.4; DB 4; Length 77536;
Best Local Similarity 49.1%; Pred. No. 0.72;
Matches 112; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 34 CAAGACGGCGGGCGCTGCTGGAGGGGTGAGCCCTGCGGACGTGTCTCGTGGCGAGAC 93
Db 31001 CGACGGGCGCGCGCGGACAGACGGCGGTGCGGACCGCGGCTGACCGCGGGTCCAC 30942

QY 94 GGTGCTTCATCAAGACCATCCGGCTCGTGGGGCGAGTGTGCTCCGCGGCCACCGGGGA 153
Db 30941 CTGCGCGGACCGCGGACCTGTGTCTACCGGCGGGTGCAGCGCGCGATCCGCTGGC 30882

QY 154 GCGCGATGGCCCTGAAGGGAGCGCTGCCCGCGCTGCCCGAGAGTGGCTTCGA 213
Db 30881 GCACTCCCTGACGGCTCGCGGTGTGCGCGCGCGGCTCTCGATCTGCGCATCG 30822

QY 214 GCGGAGGTGGAGTACAAAGGGGGCTTCCACCTGGGCATCGACGTGA 261
Db 30821 GCGGCGGACGAGCGGCTGCGGCGCCCTCGACAGTTTACCGTGA 30774

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 12:45:10 ; Search time 248 Seconds
(without alignments)
4470.886 Million cell updates/sec

Title: US-10-051-769-2

Perfect score: 261

Sequence: 1 gatcaagtgaggttcgagg.....caccctggccatcgacgtgga 261

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N Geneseq 29Jan04: *
1: geneseqn1980s: *
2: geneseqn1990s: *
3: geneseqn2000s: *
4: geneseqn2001as: *
5: geneseqn2001bs: *
6: geneseqn2002s: *
7: geneseqn2003as: *
8: geneseqn2003bs: *
9: geneseqn2003cs: *
10: geneseqn2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	261	100.0	261	7	Abx95679 Human Gli
2	261	100.0	261	9	Aad61560 Human Gli
3	261	100.0	873	7	Abx95680 Human Gli
4	261	100.0	960	7	Abx95678 Human gen
5	261	100.0	960	9	Aad61559 Human Gli
6	261	100.0	1092	9	Ade07789 Novel cod
7	261	100.0	2639	9	Adb62530 Human cdn
8	261	100.0	3832	9	Aad61564 Human Gli
9	261	100.0	4450	4	Aak51608 Human pol
10	261	100.0	4470	4	ABA09174 Human sec
11	261	100.0	4470	4	Aak52592 Human pol
12	261	100.0	4470	9	Ade09891 Novel DNA
13	261	100.0	4801	9	Adk49052 Human NOV
14	47	18.0	3847	6	Abk87764 Human cdn
15	43	16.5	263	7	Abx95681 Rat GliTE
16	43	16.5	263	9	Aad61561 Human Gli
17	22	8.4	22	7	Abx95682 Human Gli
18	22	8.4	22	9	Aad61562 Human Gli
19	21	8.0	1425	2	Aaq91639 Human son
20	21	8.0	1425	2	Aax16187 Human Shh
21	21	8.0	1425	2	Aax25622 Human son
22	21	8.0	1425	2	Aax07276 Human son
23	21	8.0	1425	2	Aax25103 Human son

24	21	8.0	1425	3	AAA30279	Partial h
25	21	8.0	1425	3	AAZ52262	Human son
26	21	8.0	1425	3	AAa27881	Human son
27	21	8.0	1425	3	AAA50451	Human son
28	21	8.0	1425	4	Aah28451	Nucleotid
29	21	8.0	1425	4	Aaf27018	Human son
30	21	8.0	1425	4	Aad10151	Human son
31	21	8.0	1425	4	Aah76112	Human son
32	21	8.0	1425	5	AAI66776	Human son
33	21	8.0	1425	5	AAC87079	Nucleotid
34	21	8.0	1425	6	ABN87549	Human son
35	21	8.0	1425	6	AAD23804	Human son
36	21	8.0	1425	6	AAK99685	cdNA enco
37	21	8.0	1425	6	ABK88636	cdNA enco
38	21	8.0	1425	8	ADA26295	Human son
39	21	8.0	1425	9	ADD25315	Human son
40	21	8.0	1425	9	AAD62097	Human son
41	21	8.0	1425	9	ADD71376	cdNA enco
42	21	8.0	1576	2	AAV18403	Human mut
43	21	8.0	1576	2	AAV18404	Human mut
44	21	8.0	1576	4	AAF84005	Human son
45	21	8.0	2274	8	ADA50078	Protein k

ALIGNMENTS

RESULT 1
ABX95679
ID ABX95679 standard; cdNA; 261 BP.
XX
AC ABX95679;
XX
DT 30-JUN-2003 (first entry)
XX
DE Human GliTEN expressed sequence tag.
XX
KW Human; ss; EST; chromosome 10q25; glioblastoma multiforme; brain cancer;
KW GliTEN; GMB; expressed sequence tag.
XX
OS Homo sapiens.
XX
PN US2003044811-A1.
XX
PD 06-MAR-2003.
XX
PF 20-OCT-2001; 2001US-00051769.
XX
PR 20-OCT-2000; 2000US-0242160P.
XX
PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX
PI McKinnon RD;
XX
DR WPI; 2003-418995/39.
XX
PT New isolated nucleic acid representing a gene product associated with
PT Glioblastoma Multiforme, designated as GliTEN, useful as a molecular
marker for tumor cell identification and classification.
XX
PS Claim 1; Page 7; 11pp; English.
XX
CC The invention relates to an isolated nucleic acid representing a gene
CC product associated with Glioblastoma Multiforme, designated as GliTEN,
CC comprising the human EST (expressed sequence tag) appearing as ABX95678,
CC or a sequence that hybridises under stringent conditions to the EST, or
CC its complement. Also included are a probe for use in identifying a
CC patient at risk for progression into the malignant phenotype, comprising
CC the nucleic acids detailed above, detecting whether a patient is at risk
CC for progression into Glioblastoma Multiforme (comprising: (a) providing a
CC sample from a patient; (b) adding the probe to the sample or performing a
CC PCR analysis using the primers appearing as ABX95682 and ABX95683; (c)
CC analysing levels of mRNA bound with the probe; and (d) treating a control

CC sample to assess the level of mRNA in a control sample, where the
CC presence of increased levels of mRNA expression in the sample in an
CC amount higher than the control sample indicates risk for progression into
CC Glioblastoma Multiforme), and a kit for use in detecting whether a
CC patient is at risk for progression into Glioblastoma Multiforme
CC (comprising the probe or primers). The probe or the kit is useful for
CC detecting whether a patient is at risk for progression into Glioblastoma
CC Multiforme (a type of brain cancer). The EST useful as molecular marker
CC for tumour cell identification and classification and for diagnosing or
CC identifying candidates at risk for progression into a malignant phenotype
CC especially in brain cancer therapy, treatment and diagnosis. The present
CC sequence is the human GLI1EN expressed sequence tag
XX
SQ Sequence 261 BP; 40 A; 81 C; 102 G; 38 T; 0 U; 0 Other;

Query Match 100.0%; Score 261; DB 7; Length 261;
Best Local Similarity 100.0%; Pred. No. 1.2e-111;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 60
DB 1 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 60
QY 61 GCTGAGCCTGCGGACGCTGTTCTTCTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
DB 61 GCTGAGCCTGCGGACGCTGTTCTTCTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
QY 121 CGTGGCGCCAGTGTGTCCTCGCCCTCGCCACCGGGAGCGCCGATGCGCCCTGAAGGGGGCTT 180
DB 121 CGTGGCGCCAGTGTGTCCTCGCCCTCGCCACCGGGAGCGCCGATGCGCCCTGAAGGGGGCTT 180
QY 181 GCCCGCCGCTGCCCCGAGGAGCTGGCTTCGAGCGGAGCTGGAGTACAAACGGGGGCTT 240
DB 181 GCCCGCCGCTGCCCCGAGGAGCTGGCTTCGAGCGGAGCTGGAGTACAAACGGGGGCTT 240
QY 241 CCACCTGGCCATCGACGTGGA 261
DB 241 CCACCTGGCCATCGACGTGGA 261

RESULT 2
AAD61560
ID AAD61560 standard; cDNA; 261 BP.
XX
AC AAD61560;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human GLI1EN EST clone cDNA #1.
XX
KW Human; glioblastoma multiforme; GBM; GLI1EN; brain cancer; diagnosis;
KW Gene therapy; expressed sequence tag; ss.
XX
OS Homo sapiens.
XX
PN US2003108915-A1.
XX
PD 12-JUN-2003.
XX
PF 20-AUG-2002; 2002US-00224624.
XX
PR 20-OCT-2000; 2000US-0242160P.
PR 20-OCT-2001; 2001US-00051769.
XX
PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX
PI McKinnon RD;
XX
DR WPI; 2003-810848/76.
XX
PT Novel glioblastoma multiforme associated protein GLI1EN useful for
PT treating glioblastoma multiforme and diagnosing brain cancer.
XX

PS Claim 38; Page 11; Opp; English.

XX The present invention provides novel glioblastoma multiforme (GBM)
CC associated protein GLI1EN useful for treating glioblastoma multiforme and
CC diagnosing brain cancer. The invention is useful in brain cancer therapy,
CC treatment and diagnosis. The invention is also useful in gene therapy.
CC The present sequence is human GLI1EN EST clone cDNA. This EST clone is
CC used in the exemplification of the invention
XX

SQ Sequence 261 BP; 40 A; 81 C; 102 G; 38 T; 0 U; 0 Other;

Query Match 100.0%; Score 261; DB 9; Length 261;
Best Local Similarity 100.0%; Pred. No. 1.2e-111;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 60
DB 1 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 60
QY 61 GCTGAGCCTGCGGACGCTGTTCTTCTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
DB 61 GCTGAGCCTGCGGACGCTGTTCTTCTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
QY 121 CGTGGCGCCAGTGTGTCCTCGCCCTCGCCACCGGGAGCGCCGATGCGCCCTGAAGGGGGCTT 180
DB 121 CGTGGCGCCAGTGTGTCCTCGCCCTCGCCACCGGGAGCGCCGATGCGCCCTGAAGGGGGCTT 180
QY 181 GCCCGCCGCTGCCCCGAGGAGCTGGCTTCGAGCGGAGCTGGAGTACAAACGGGGGCTT 240
DB 181 GCCCGCCGCTGCCCCGAGGAGCTGGCTTCGAGCGGAGCTGGAGTACAAACGGGGGCTT 240
QY 241 CCACCTGGCCATCGACGTGGA 261
DB 241 CCACCTGGCCATCGACGTGGA 261

RESULT 3
ABX95680
ID ABX95680 standard; DNA; 873 BP.
XX
AC ABX95680;
XX
DT 30-JUN-2003 (first entry)
XX
DE Human GLI1EN open reading frame.
XX
KW Human; ds; chromosome 10q25; glioblastoma multiforme; brain cancer;
KW GLI1EN; GMB.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..873
FT /tag= a
FT /partial
FT /product= "GLI1EN"
FT /note= "No stop codon shown"
XX
PN US2003044811-A1.
XX
PD 06-MAR-2003.
XX
PF 20-OCT-2001; 2001US-00051769.
XX
PR 20-OCT-2000; 2000US-0242160P.
XX
PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX
PI McKinnon RD;
XX
DR WPI; 2003-418995/39.
XX
PT New isolated nucleic acid representing a gene product associated with

RESULT 5

AA61559
ID AAD61559 standard; cDNA; 960 BP.
AC
XX AAD61559;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human GliTEN partial cDNA.
XX
KW Human; glioblastoma multiforme; GBM; GliTEN; brain cancer; diagnosis;
KW gene therapy; chromosome 10; gene; ss.
XX
OS Homo sapiens.
XX
FN US2003108915-A1.
XX
PD 12-JUN-2003.
XX
PF 20-AUG-2002; 2002US-00224624.
XX
PR 20-OCT-2000; 2000US-0242160P.
XX
PR 20-OCT-2001; 2001US-00051769.
XX
PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX
PI McKinnon RD;
XX
DR WPI; 2003-810848/76.
XX
PT Novel glioblastoma multiforme associated protein GliTEN useful for
PT treating glioblastoma multiforme and diagnosing brain cancer.
XX
PS Example 2; Fig 1; Opp; English.
XX
CC The present invention provides novel glioblastoma multiforme (GBM)
CC associated protein GliTEN useful for treating glioblastoma multiforme and
CC diagnosing brain cancer. The invention is useful in brain cancer therapy,
CC treatment and diagnosis. The invention is also useful in gene therapy.
CC The present sequence is human GliTEN partial cDNA. The human GliTEN gene
CC is located on chromosome 10. This partial cDNA is used in the
CC exemplification of the invention
XX
SQ Sequence 960 BP; 146 A; 324 C; 329 G; 161 T; 0 U; 0 Other;
Query Match 100.0%; Score 261; DB 9; Length 960;
Best Local Similarity 100.0%; Pred. No. 1.2e-111;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCAAGGTGAGTTTCGAGGAGCTCTCGACACCAAGACCGCGCGCGCTGCTGGAGGG 60
DB 396 GATCAAGGTGAGTTTCGAGGAGCTCTCGACACCAAGACCGCGCGCGCTGCTGGAGGG 455
QY 61 GCTGAGCTCGGGACGTTCCTCGGGGAGACGCTGCTTCATCAAGACCATCCGGCT 120
DB 456 GCTGAGCTCGGGACGTTCCTCGGGGAGACGCTGCTTCATCAAGACCATCCGGCT 515
QY 121 CGTGGGCGAGTCTGTCGCCCTCGGCACCGGGGAGCCCGATGCGCCCTGAAGGGGAGCGCT 180
DB 516 CGTGGGCGAGTCTGTCGCCCTCGGCACCGGGGAGCCCGATGCGCCCTGAAGGGGAGCGCT 575
QY 181 GCCCGCGCCCTGCCCGGAGGAGCTGGCTTCGAGCGGAGGTGGAGTACAACGGGGGCTT 240
DB 576 GCCCGCGCCCTGCCCGGAGGAGCTGGCTTCGAGCGGAGGTGGAGTACAACGGGGGCTT 635
QY 241 CCACCTGGCCATCGACGTGGA 261
DB 636 CCACCTGGCCATCGACGTGGA 656

RESULT 6

ADE07789
ID ADE07789 standard; DNA; 1092 BP.
XX
AC ADE07789;
XX
DT 29-JAN-2004 (first entry)
XX
DE Novel coding sequence (useful for identifying genetic disorders) #855.
XX
KW novel gene; novel protein; tissue marker; molecular weight marker;
KW chromosome marker; genetic disorder; gene; ds.
XX
OS Unidentified.
XX
PN WO2003054152-A2.
XX
PD 03-JUL-2003.
XX
PF 10-DEC-2002; 2002WO-US039555.
XX
PR 10-DEC-2001; 2001US-0339739P.
XX
PR 11-DEC-2001; 2001US-0339453P.
XX
PR 14-MAR-2002; 2002US-0365091P.
XX
PR 14-MAR-2002; 2002US-0365384P.
XX
PR 12-APR-2002; 2002US-0372381P.
XX
PR 12-APR-2002; 2002US-0372615P.
XX
PR 22-APR-2002; 2002US-00128558.
XX
PR 24-APR-2002; 2002US-0376045P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
PI Choeh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;
PI Ma Y, Wang D, Chen R, Xu C, Boyle BU;
XX
DR WPI; 2003-569235/53.
XX
DR P-PSDB; ADE08700.
XX
PT New polynucleotides, useful for expressing recombinant proteins for
PT analysis, characterization or therapeutic use, or as markers for tissues
PT in which the corresponding protein is preferentially expressed.
XX
PS Claim 1; SEQ ID NO 855; 1177pp; English.
XX
CC The invention comprises the amino acid and coding sequences of novel
CC proteins. The DNA and protein sequences of the invention are useful as:
CC markers for tissues in which the corresponding protein is preferentially
CC expressed; as molecular weight markers on gels; as chromosome markers or
CC tags; to identify chromosomes or to map related gene positions; and to
CC compare with endogenous DNA sequences in patients to identify potential
CC genetic disorders. The present DNA sequence represents a gene of the
CC invention.
XX
SQ Sequence 1092 BP; 190 A; 377 C; 322 G; 203 T; 0 U; 0 Other;
Query Match 100.0%; Score 261; DB 9; Length 1092;
Best Local Similarity 100.0%; Pred. No. 1.1e-111;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCAAGGTGAGTTTCGAGGAGCTGCTCGACACCAAGACCGCGCGCGCTGCTGGAGGG 60
DB 366 GATCAAGGTGAGTTTCGAGGAGCTGCTCGACACCAAGACCGCGCGCGCTGCTGGAGGG 425
QY 61 GCTGAGCTCGGGACGTTCCTCGGGGAGACGCTGCTTCATCAAGACCATCCGGCT 120
DB 426 GCTGAGCTCGGGACGTTCCTCGGGGAGACGCTGCTTCATCAAGACCATCCGGCT 485
QY 121 CGTGGGCGAGTCTGTCGCCCTCGGCACCGGGGAGCCCGATGCGCCCTGAAGGGGAGCGCT 180
DB 486 CGTGGGCGAGTCTGTCGCCCTCGGCACCGGGGAGCCCGATGCGCCCTGAAGGGGAGCGCT 545
QY 181 GCCCGCGCCCTGCCCGGAGGAGCTGGCTTCGAGCGGAGGTGGAGTACAACGGGGGCTT 240

Db 546 GCCCGCGCTGCCCGAGAGCTGGCTTCGAGCGGAGGTGAGTACAAACGGGGCTT 605
QY 241 CCACCTGGCCATCGAGTGA 261
Db ||||||||||||||||
606 CCACCTGGCCATCGAGTGA 626

RESULT 7
ADB62530
ID ADB62530 standard; cDNA; 2639 BP.
XX AC ADB62530;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human cDNA encoding clone HHDPC2008160.
XX
KW Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW tissue regeneration; cell regeneration; membrane protein;
KW signal transduction-related protein; transcription-related protein;
KW osteoporosis; neurological disease; cancer; tumour.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 214..2639
FT /tag= a
FT /partial
FT /product= "Clone HHDPC2008160 protein"
XX
PN EP1308459-A2.
XX
PD 07-MAY-2003.
XX
XX 28-MAR-2002; 2002EP-00007401.
XX
PR 05-NOV-2001; 2001JP-00379298.
PR 25-JAN-2002; 2002US-00350978.
XX
XX (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamanoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
XX WPI; 2003-450961/43.
DR P-PSDB; ADB64500.
XX
XX New polynucleotides and polypeptides, useful for developing a diagnostic
PT marker or medicines for regulation of their expression and activity, or
PT as targets of gene therapy.
XX
XX Claim 1; Page; 222pp; English.
XX
XX The invention discloses a polynucleotide comprising a sequence selected
CC from 1970 fully defined nucleotide sequences which encode novel
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC or its partial peptide, an antibody binding to the polypeptide or peptide
CC of the polynucleotide, immunologically assaying the polypeptide or
CC peptide of the polynucleotide by contacting the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesising the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotides and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease-related proteins and genes
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
XX

CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a cDNA of the invention. Note: Some of the sequence
CC data for this patent is not represented in the printed specification, but
CC is based on sequence information supplied by the European Patent Office.
XX
SQ Sequence 2639 BP; 657 A; 702 C; 719 G; 561 T; 0 U; 0 Other;

Query Match 100.0%; Score 261; DB 9; Length 2639;
Best Local Similarity 100.0%; Pred. No. 1.1e-111;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGTGGAGTTCGAGGAGCTGTCGAGACCAAGACGCGCGCGCTGGAGGG 60
Db ||||||||||||||||
579 GATCAAGTGGAGTTCGAGGAGCTGTCGAGACCAAGACGCGCGCGCTGGAGGG 638

QY 61 GCTGAGCCTGCGGAGCTGTTCTTGGGCGAGACGCTGCCCTTCATCAAGACCATCCGGCT 120
Db ||||||||||||||||
639 GCTGAGCCTGCGGAGCTGTTCTTGGGCGAGACGCTGCCCTTCATCAAGACCATCCGGCT 698

QY 121 CGTGGCGCCAGTCTGTCCTCGGCCACCGGGGAGCCGATGGCCCTGAAGGGGAGGGCT 180
Db ||||||||||||||||
699 CGTGGCGCCAGTCTGTCCTCGGCCACCGGGGAGCCGATGGCCCTGAAGGGGAGGGCT 758

QY 181 GCCGCGCGCTGCCGAGGAGCTGGCTTCGAGGCGGAGGTGAGTACAAACGGGGCTT 240
Db ||||||||||||||||
759 GCCGCGCGCTGCCGAGGAGCTGGCTTCGAGGCGGAGGTGAGTACAAACGGGGCTT 818

QY 241 CCACCTGGCCATCGAGTGA 261
Db ||||||||||||||||
819 CCACCTGGCCATCGAGTGA 839

RESULT 8
AAD61564
ID AAD61564 standard; cDNA; 3832 BP.
XX
AC AAD61564;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human GLI1EN cDNA.
XX
XX Human; glioblastoma multiforme; GBM; GLI1EN; brain cancer; diagnosis;
XX gene therapy; chromosome 10; gene; ss.
XX
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 178..3642
FT /tag= a
FT /product= "Human GLI1EN protein"
FT /note= "The CDS is specifically claimed in claim 1"
FT primer_bind 1011..1032
FT /tag= b
FT /bound_moiety= "5' primer #843"
FT primer_bind 1405..1429
FT /tag= c
FT /bound_moiety= "5' primer #1405"
FT primer_bind 2083..2108
FT /tag= d
FT /bound_moiety= "5' primer #2083"
FT primer_bind 2356..2378
FT /tag= e
FT /bound_moiety= "5' primer #2355"
FT primer_bind 3625..3650
FT /tag= f
FT /bound_moiety= "3' primer #3636"
XX
XX US2003108915-A1.
PN
XX 12-JUN-2003.
PD
XX

DE Human secreted protein homologue-encoding cDNA, SEQ ID NO:950.
XX
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytostatic; osteopathic; vasotrophic; cardiant; virucide; antibacterial;
KW antifungal; vulnery; antiulcer; ss.
OS Homo sapiens.
XX
XX WO200157188-A2.
XX
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US003800.
XX
XX 03-FEB-2000; 2000US-00496914.
XX 27-APR-2000; 2000US-00560875.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-457740/49.
XX P-PSDB; ABB11930.
XX
XX
XX Human proteins and DNA encoding sequences useful for preventing, treating
XX or ameliorating a medical condition in a mammalian subject e.g. arthritis
XX and cancer.
XX
XX Claim 1; Page 813-814; 1963pp; English.
XX
XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
XX sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
XX invention also relates to vectors and recombinant host cells comprising a
XX nucleotide of the invention, methods of producing the novel polypeptides,
XX antibodies against the polypeptides, methods of detecting the nucleotides
XX or polypeptides in a sample, and methods of identifying compounds which
XX bind to polypeptides of the invention. Although novel, many of the
XX polypeptides of the invention have homology to known proteins, thereby
XX giving an insight into their probable biological activities, and hence
XX potential therapeutic applications. The polypeptides of the invention may
XX have various activities, including cytokine, cell proliferation or cell
XX differentiation activities; stem cell growth factor activity;
XX haematopoiesis regulatory activity; activin- or inhibin-related activities;
XX immunomodulatory activity; activin- or inhibin-related activities;
XX chemotactic or chemokinetic activities; haemostatic, thrombotic or
XX thrombolytic activities; receptor or ligand activities; or may be
XX involved in oncogenesis, cancer cell proliferation or metastasis.
XX Depending on their biological activities, polypeptides and nucleotides of
XX the invention are useful for preventing, treating or ameliorating medical
XX conditions, e.g., by protein or gene therapy. Such conditions include
XX cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
XX disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
XX proliferative retinopathy, atherosclerosis, coronary heart disease,
XX arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
XX vascular growth. Polypeptides involved with tissue regeneration and
XX repair (or nucleic acids encoding them) may be used to promote wound
XX healing (e.g., of burns, incisions and ulcers), while those with
XX immunomodulatory activities may be used in the treatment of viral,
XX bacterial and fungal infections in addition to immune disorders.
XX Polypeptides with growth factor activity may be used in cell cultures to
XX promote cell growth. For example, such polypeptides may be used to
XX manipulate stem cells in culture to give rise to neuroepithelial cells
XX that can be used to augment or replace cells damaged by illness,

CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a cDNA encoding a
CC novel human polypeptide of the invention
XX
SQ Sequence 4470 BP; 1254 A; 991 C; 1054 G; 1171 T; 0 U; 0 Other;
Query Match 100.0%; Score 261; DB 4; Length 4470;
Best Local Similarity 100.0%; Pred. No. 1.1e-111;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCAAGGTGGAGTTCGAGCAGCTGCTGCAGACCAAGACGGCGCGGCTGCTGAGGG 60
DB 366 GATCAAGGTGGAGTTCGAGCAGCTGCTGCAGACCAAGACGGCGCGGCTGCTGAGGG 425
QY 61 GCTGAGCCCTGCGGACGCTTCTCTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
DB 426 GCTGAGCCCTGCGGACGCTTCTCTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 485
QY 121 CGTGGCGCCAGTCTGCTCCCTCGGCACCGGAGCCCGATGCGCCCTGAAGGGGAGCGCT 180
DB 486 CGTGGCGCCAGTCTGCTCCCTCGGCACCGGAGCCCGATGCGCCCTGAAGGGGAGCGCT 545
QY 181 GCCCGCGCGCTGCCCGAGGAGCTGGCCCTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 240
DB 546 GCCCGCGCGCTGCCCGAGGAGCTGGCCCTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 605
QY 241 CCACCTGGCCATCGACGTGGA 261
DB 606 CCACCTGGCCATCGACGTGGA 626
RESULT 11
AAK52592
ID AAK52592 standard; cDNA; 4470 BP.
XX
AC AAK52592;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 2121.
DE
DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
XX WO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US004098.
XX
XX 03-FEB-2000; 2000US-00496914.
XX 27-APR-2000; 2000US-00560875.
XX 20-JUN-2000; 2000US-00598075.
XX 19-JUL-2000; 2000US-00620325.
XX 01-SEP-2000; 2000US-00654936.
XX 15-SEP-2000; 2000US-00663561.
XX 20-OCT-2000; 2000US-00693325.
XX 30-NOV-2000; 2000US-00728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
XX Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI; 2001-476283/51.
XX P-PSDB; AAM79459.
XX

PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.

XX Claim 1; Page 4496-4497; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication

XX Sequence 4470 BP; 1254 A; 991 C; 1054 G; 1171 T; 0 U; 0 Other;

Query Match 100.0%; Score 261; DB 4; Length 4470;

Best Local Similarity 100.0%; Pred. No. 1.1e-111;

Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGGTGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCGCTGCTGGAGGG 60

DB 366 GATCAAGGTGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCGCTGCTGGAGGG 425

QY 61 GCTGAGCTCGGGGACGTGTTCTTGGGGGAGACGGTGCCTTCAACAAGACCATCCGGCT 120

DB 426 GCTGAGCTCGGGGACGTGTTCTTGGGGGAGACGGTGCCTTCAACAAGACCATCCGGCT 485

QY 121 CGTGGGCGCACTGTCGCCCTCGCCACCGGGGAGCCCGATGCCCTGAAGGGAGCGCT 180

DB 486 CGTGGGCGCACTGTCGCCCTCGCCACCGGGGAGCCCGATGCCCTGAAGGGAGCGCT 545

QY 181 GCCCGCGCCCTGCCCCGAGGAGCTGGCCCTTCGAGCGGAGTGGAGTACAACGGGGGCTT 240

DB 546 GCCCGCGCCCTGCCCCGAGGAGTGGCCCTTCGAGCGGAGTGGAGTACAACGGGGGCTT 605

QY 241 CCACCTGGCCATCGACGTGGA 261

DB 606 CCACCTGGCCATCGACGTGGA 626

RESULT 12

AD809891

XX ADE09891 standard; DNA; 4470 BP.

XX AC ADE09891;

XX XX 29-JAN-2004 (first entry)

XX Novel DNA-related contig nucleotide sequence #613.

DE novel gene; novel protein; tissue marker; molecular weight marker;

XX chromosome marker; genetic disorder; contig; ds.

XX Unidentified.

XX WO2003054152-A2.

XX 03-JUL-2003.

XX 10-DEC-2002; 2002WO-US039555.

XX 10-DEC-2001; 2001US-0339739P.

XX 11-DEC-2001; 2001US-0339453P.

XX 14-MAR-2002; 2002US-0365091P.

XX 14-MAR-2002; 2002US-0365384P.

XX 12-APR-2002; 2002US-0372381P.

XX 12-APR-2002; 2002US-0372615P.

PR 22-APR-2002; 2002US-00128558.

PR 24-APR-2002; 2002US-0376045P.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;

PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;

PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;

XX WPI; 2003-569235/53.

DR New polynucleotides, useful for expressing recombinant proteins for

XX analysis, characterization or therapeutic use, or as markers for tissues

PT in which the corresponding protein is preferentially expressed.

PT Disclosure; SEQ ID NO 2435; 1177pp; English.

XX The invention comprises the amino acid and coding sequences of novel

CC proteins. The DNA and protein sequences of the invention are useful as:

CC markers for tissues in which the corresponding protein is preferentially

CC expressed; as molecular weight markers on gels; as chromosome markers or

CC tags; to identify chromosomes or to map related gene positions; and to

CC compare with endogenous DNA sequences in patients to identify potential

CC genetic disorders. The present DNA sequence was used in the

CC exemplification of the invention.

XX Sequence 4470 BP; 1254 A; 991 C; 1054 G; 1171 T; 0 U; 0 Other;

QY Query Match 100.0%; Score 261; DB 9; Length 4470;

DB Best Local Similarity 100.0%; Pred. No. 1.1e-111;

Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGGTGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCGCTGCTGGAGGG 60

DB 366 GATCAAGGTGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCGCTGCTGGAGGG 425

QY 61 GCTGAGCTCGGGGACGTGTTCTTGGGGGAGACGGTGCCTTCAACAAGACCATCCGGCT 120

DB 426 GCTGAGCTCGGGGACGTGTTCTTGGGGGAGACGGTGCCTTCAACAAGACCATCCGGCT 485

QY 121 CGTGGGCGCACTGTCGCCCTCGCCACCGGGGAGCCCGATGCCCTGAAGGGAGCGCT 180

DB 486 CGTGGGCGCACTGTCGCCCTCGCCACCGGGGAGCCCGATGCCCTGAAGGGAGCGCT 545

QY 181 GCCCGCGCCCTGCCCCGAGGAGCTGGCCCTTCGAGCGGAGTGGAGTACAACGGGGGCTT 240

DB 546 GCCCGCGCCCTGCCCCGAGGAGTGGCCCTTCGAGCGGAGTGGAGTACAACGGGGGCTT 605

QY 241 CCACCTGGCCATCGACGTGGA 261

DB 606 CCACCTGGCCATCGACGTGGA 626

RESULT 13

AD49052

ID ADD49052 standard; DNA; 4801 BP.

XX AC ADD49052;

XX 15-JAN-2004 (first entry)

XX Human NOV6a coding sequence, SEQ ID 25.

XX Antidiabetic; anorectic; cardiant; hypotensive; antiarteriosclerotic;

XX virucide; antibacterial; fungicide; protozoacide; nootropic;

XX neuroprotective; antiparkinsonian; anticonvulsant; osteopathic;

XX antiarthritis; antinflammatory; dermatological; antiasthmatic;

XX antileptic; gene therapy; NOV protein; metabolic disorder; diabetes;

XX obesity; viral infection; bacterial infection; fungal infection;

XX helminthic infection; protozoal infection; anorexia; cancer;

XX cardiovascular disease; hypertension; atherosclerosis;

XX neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;

XX epilepsy; immune disorder; osteoarthritis; haematopoietic disorder;

CC protein or peptide, an non-human animal which has deletion of the gene
CC function encoding the protein or its partial peptide on the chromosome,
CC or has overexpression of the protein or its partial peptide and screening
CC substances promoting or inhibiting immune induction activity by using the
CC protein or its partial peptide, the test substances and T cells, and
CC measuring and evaluating immune induction activity in T cells. The
CC antigen and DNA encoding it, are applicable in diagnosis and treatment of
CC (e.g. by peptide or gene therapy) glioma, human malignant brain tumour
CC and other cancers, and for studying onset of glioma. The present sequence
CC encodes the glioma antigen KU-GB-5

SQ Sequence 3847 BP; 1197 A; 744 C; 819 G; 1086 T; 0 U; 1 Other;

Query Match 18.0%; Score 47; DB 6; Length 3847;
Best Local Similarity 100.0%; Pred. No. 3.7e-12; Length 3847;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 GCGGAGGTGGAGTACAAACGGGGGCTTCCACCTGGCGCATCGACGTGGA 261

Db 1 GCGGAGGTGGAGTACAAACGGGGGCTTCCACCTGGCGCATCGACGTGGA 47

RESULT 15

ABX95681

ID ABX95681 standard; cDNA; 263 BP.

AC AC

AC ABX95681;

XX DT 30-JUN-2003 (first entry)

XX DE Rat GliTEN expressed sequence tag.

XX KW Rat; ss; EST; glioblastoma multiforme; brain cancer; GliTEN; GMB;

XX KW expressed sequence tag.

XX OS Rattus Norvegicus.

XX PN US2003044811-A1.

XX PD 06-MAR-2003.

XX PF 20-OCT-2001; 2001US-00051769.

XX PR 20-OCT-2000; 2000US-0242160P.

XX PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

XX PI McKinnon RD;

XX PS WPI; 2003-418995/39.

XX PT New isolated nucleic acid representing a gene product associated with

XX PT Glioblastoma Multiforme, designated as GliTEN, useful as a molecular

XX PT marker for tumor cell identification and classification.

XX PS Example 2; Page 4; 11pp; English.

XX CC The invention relates to an isolated nucleic acid representing a gene

XX CC product associated with Glioblastoma Multiforme, designated as GliTEN,

XX CC comprising the human EST (expressed sequence tag) appearing as ABX95678,

XX CC or a sequence that hybridises under stringent conditions to the EST, or

XX CC its complement. Also included are a probe for use in identifying a

XX CC patient at risk for progression into the malignant phenotype, comprising

XX CC the nucleic acids detailed above, detecting whether a patient is at risk

XX CC for progression into Glioblastoma Multiforme (comprising: (a) providing a

XX CC sample from a patient; (b) adding the probe to the sample or performing

XX CC PCR analysis using the primers appearing as ABX95682 and ABX95683; (c)

XX CC analysing levels of mRNA bound with the probe; and (d) treating a control

XX CC sample to assess the level of mRNA in a control sample, where the

XX CC presence of increased levels of mRNA expression in the sample in an

XX CC amount higher than the control sample indicates risk for progression into

XX CC Glioblastoma Multiforme), and a kit for use in detecting whether a

XX CC patient is at risk for progression into Glioblastoma Multiforme

CC (comprising the probe or primers). The probe or the kit is useful for
CC detecting whether a patient is at risk for progression into Glioblastoma
CC Multiforme (a type of brain cancer). The EST useful as molecular marker
CC for tumour cell identification and classification and for diagnosing or
CC identifying candidates at risk for progression into a malignant phenotype
CC especially in brain cancer therapy, treatment and diagnosis. The present
CC sequence is a Rat GliTEN expressed sequence tag

XX SQ Sequence 263 BP; 41 A; 78 C; 94 G; 50 T; 0 U; 0 Other;

Query Match 16.5%; Score 43; DB 7; Length 263;
Best Local Similarity 100.0%; Pred. No. 3.1e-10;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGGTGGAGTTCGAGGAGCTCTGCAGACCAAGACGGCC 43

Db 1 GATCAAGGTGGAGTTCGAGGAGCTCTGCAGACCAAGACGGCC 43

Search completed: July 18, 2004, 14:51:45

Job time : 250 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: July 18, 2004, 13:33:30 ; Search time 1612 Seconds
(without alignments)
7017.700 Million cell updates/sec

Title: US-10-051-769-2
Perfect score: 261
Sequence: 1 gatcaagtgaggcttgagg.....cactggccatgcagctgga 261

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues
Word size : 0
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- GenEmbl:*
- 1: gb.ba.*
 - 2: gb.htg.*
 - 3: gb.in.*
 - 4: gb.om.*
 - 5: gb.ov.*
 - 6: gb.pat.*
 - 7: gb.ph.*
 - 8: gb.pl.*
 - 9: gb.pr.*
 - 10: gb.ro.*
 - 11: gb.sts.*
 - 12: gb.sy.*
 - 13: gb.un.*
 - 14: gb.vi.*
 - 15: em.ba.*
 - 16: em.fun.*
 - 17: em.hum.*
 - 18: em.in.*
 - 19: em.mu.*
 - 20: em.om.*
 - 21: em.or.*
 - 22: em.ov.*
 - 23: em.pat.*
 - 24: em.ph.*
 - 25: em.pl.*
 - 26: em.ro.*
 - 27: em.sts.*
 - 28: em.un.*
 - 29: em.vi.*
 - 30: em.htg.hum.*
 - 31: em.htg.inv.*
 - 32: em.htg.other.*
 - 33: em.htg.mus.*
 - 34: em.htg.pln.*
 - 35: em.htg.rod.*
 - 36: em.htg.mam.*
 - 37: em.htg.vrt.*
 - 38: em.sy.*
 - 39: em.hgo.hum.*
 - 40: em.hgo.mus.*
 - 41: em.hgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	261	100.0	2623	9	BC012186	BC012186 Homo sapi
2	261	100.0	2639	6	AX747159	AX747159 Sequence
3	261	100.0	2639	9	AK091746	AK091746 Homo sapi
4	261	100.0	3856	9	BC028375	BC028375 Homo sapi
5	261	100.0	49052	9	AL359836	AL359836 Human DNA
6	261	100.0	120578	9	AC005887	AC005887 catb 173
7	47	18.0	3847	6	BD169701	BD169701 Human gl1
8	45	17.2	68196	2	AC108407	AC108407 Mus muscu
9	45	17.2	221524	10	AC139040	AC139040 Mus muscu
10	45	17.2	243412	2	AC127769	AC127769 Rattus no
11	23	8.8	179734	2	BX511030	BX511030 Danio rer
12	22	8.4	3291	1	AF076980	AF076980 Streptomy
13	21	8.0	1425	6	AR063085	AR063085 Sequence
14	21	8.0	1425	6	AR122612	AR122612 Sequence
15	21	8.0	1425	6	AR164240	AR164240 Sequence
16	21	8.0	1425	6	BD228646	BD228646 Regulatio
17	21	8.0	1425	6	BD243626	BD243626 Methods a
18	21	8.0	1425	6	BD247295	BD247295 Method fo
19	21	8.0	1425	6	AR208912	AR208912 Sequence
20	21	8.0	1425	6	AR226520	AR226520 Sequence
21	21	8.0	1425	6	AR342212	AR342212 Sequence
22	21	8.0	1425	6	AR381233	AR381233 Sequence
23	21	8.0	1425	6	AR383134	AR383134 Sequence
24	21	8.0	1425	6	AR392796	AR392796 Sequence
25	21	8.0	1425	6	AR404901	AR404901 Sequence
26	21	8.0	1425	6	AR412225	AR412225 Sequence
27	21	8.0	1425	6	AX282792	AX282792 Sequence
28	21	8.0	1425	6	AX354859	AX354859 Sequence
29	21	8.0	1425	6	AX417081	AX417081 Sequence
30	21	8.0	1425	6	AX536184	AX536184 Sequence
31	21	8.0	1425	6	BD078402	BD078402 Method of
32	21	8.0	1576	9	HUMSHH	L38518 Homo sapien
33	21	8.0	2214	10	AF325507	AF325507 Mus muscu
34	21	8.0	2235	10	AF028009	AF028009 Mus muscu
35	21	8.0	2274	6	AR282728	AR282728 Sequence
36	21	8.0	2704	6	E02148	E02148 Rat protein
37	21	8.0	2704	10	RATPKCEA	M18331 Rat protein
38	21	8.0	2707	6	A37235	A37235 Sequence 4
39	21	8.0	2750	4	RABNPKC	M20014 Rabbit prot
40	21	8.0	2850	6	AX054864	AX054864 Sequence
41	21	8.0	2850	6	AX146544	AX146544 Sequence
42	21	8.0	13414	9	AY422195	AY422195 Homo sapi
43	21	8.0	48469	9	AC002484	AC002484 Human cos
44	21	8.0	68555	2	AC104963	AC104963 Homo sapi
45	21	8.0	154737	9	AC078834	AC078834 Homo sapi

ALIGNMENTS

RESULT 1
BC012186
LOCUS BC012186 Homo sapiens, clone IMAGE:4564853, mRNA. PRI 06-AUG-2001
DEFINITION BC012186 Homo sapiens, clone IMAGE:4564853, mRNA.
ACCESSION BC012186
VERSION BC012186.1 GI:15082555
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2623) Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) National Institutes of Health, Mammalian

REMARK
COMMENT

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saedi, Jacqueline
Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Nataaja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 29 Row: h Column: 17

This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, GenomeScan gene prediction.

FEATURES

source
1. .2623
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4564853"
/tissue_type="kidney, renal cell adenocarcinoma"
/clone_lib="NIH MGC 14"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

ORIGIN

Query Match 100.0%; Score 261; DB 9; Length 2623;
Best Local Similarity 100.0%; Pred. No. 1e-120;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCGGCGCTGCGAGGG 60
|||||
Db 543 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCGGCGCTGCGAGGG 602
|||||
Qy 61 GCTGAGCCTGGCGGACGCTGTTCTCTGGGCGAGACGCTGCTTCATCAAGACCATCCGGCT 120
|||||
Db 603 GCTGAGCCTGGCGGACGCTGTTCTCTGGGCGAGACGCTGCTTCATCAAGACCATCCGGCT 662
|||||
Qy 121 GCTCGGGCGAGTGTGCTCGGCTCGGACACGGGGAGCCGATGCGCTTGAAGGGAGGCGCT 180
|||||
Db 663 GCTCGGGCGAGTGTGCTCGGCTCGGACACGGGGAGCCGATGCGCTTGAAGGGAGGCGCT 722
|||||
Qy 181 GCCCGCGCTGCGCGGAGGCTGCTTCGAGCGGAGCTGAGGTGAGTACAAACGGGGGCTT 240
|||||
Db 723 GCCCGCGCTGCGCGGAGGCTGCTTCGAGCGGAGCTGAGGTGAGTACAAACGGGGGCTT 782
|||||
Qy 241 CCACCTGGCCATCGACGTGGA 261
|||||
Db 783 CCACCTGGCCATCGACGTGGA 803
|||||

RESULT 2

AX747159
LOCUS AX747159 2639 bp mRNA linear PAT 20-JUN-2003
DEFINITION Sequence 684 from Patent EP1308459.
ACCESSION AX747159
VERSION AX747159.1 GI:32131547
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1
AUTHORS
Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,
Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,
Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and
Masuho, Y.

TITLE

Full-length cDNA sequences

JOURNAL

Patent: EP 1308459-A 684 07-MAY-2003;

JOURNAL

Helix Research Institute (JP) ; Research Association for

JOURNAL

Biotechnology (JP)

FEATURES

Location/Qualifiers

source

1. .2639

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 261; DB 6; Length 2639;
Best Local Similarity 100.0%; Pred. No. 1e-120;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCGGCGCTGCGAGGG 60
|||||
Db 579 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCGGCGCTGCGAGGG 638
|||||
Qy 61 GCTGAGCCTGGCGGACGCTGTTCTCTGGGCGAGACGCTGCTTCATCAAGACCATCCGGCT 120
|||||
Db 639 GCTGAGCCTGGCGGACGCTGTTCTCTGGGCGAGACGCTGCTTCATCAAGACCATCCGGCT 698
|||||
Qy 121 GCTCGGGCGAGTGTGCTCGGCTCGGACACGGGGAGCCGATGCGCTTGAAGGGAGGCGCT 180
|||||
Db 699 GCTCGGGCGAGTGTGCTCGGCTCGGACACGGGGAGCCGATGCGCTTGAAGGGAGGCGCT 758
|||||
Qy 181 GCCCGCGCTGCGCGGAGGCTGCTTCGAGCGGAGCTGAGGTGAGTACAAACGGGGGCTT 240
|||||
Db 759 GCCCGCGCTGCGCGGAGGCTGCTTCGAGCGGAGCTGAGGTGAGTACAAACGGGGGCTT 818
|||||
Qy 241 CCACCTGGCCATCGACGTGGA 261
|||||
Db 819 CCACCTGGCCATCGACGTGGA 839
|||||

RESULT 3

AX091746
LOCUS AX091746 2639 bp mRNA linear PRI 15-JUL-2002
DEFINITION Homo sapiens cDNA FLJ34427 fis, clone HHDPC2008816.
ACCESSION AX091746
VERSION AX091746.1 GI:21750192
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1
AUTHORS
Nishi, T., Ota, T., Nakagawa, S., Senoh, A., Mizuuchi, H., Inagaki, H.,
Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S.,
Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T.,
Kamura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M.,
Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kaneshiro, K.,
Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y.,
Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.

NEDO human cDNA sequencing project

TITLE

Unpublished

JOURNAL

2 (bases 1 to 2639)

REFERENCE

Isogai, T. and Yamamoto, J.

AUTHORS

Direct Submission

TITLE

Submitted (04-JUL-2002)

JOURNAL

Takao Isogai, FLJ Project (HRI Team); 2-6-7

Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan

(E-mail: genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)

NEDO human cDNA sequencing project supported by Ministry of

Economy, Trade and Industry of Japan; cDNA full insert sequencing:

Research Association for Biotechnology (RAB); cDNA library


```
REFERENCE 4 Street, Waltham, MA 02154, USA
AUTHORS (bases 1 to 120578)
TITLE Direct Submission
JOURNAL Submitted (02-MAR-1999) Genome Therapeutics Corporation, 100 Beaver
REFERENCE 5 Street, Waltham, MA 02154, USA
AUTHORS (bases 1 to 120578)
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver
REMARK Street, Waltham, MA 02154, USA
COMMENT Vector Sequence Clipped
FEATURES On Nov 5, 1999 this sequence version replaced gi:4314331.
source Location/Qualifiers
1..120578
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/map="10q25"
/clone_lib="CIT987SK-1173112"
ORIGIN
Query Match 100.0%; Score 261; DB 9; Length 120578;
Best Local Similarity 100.0%; Pred. No. 7.8e-121;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCAAGTGGAGTTCGAGAGCTGCTGCAGACCAAGACGCGCGGCGCTGCTGGAGGG 54035
Db GATCAAGTGGAGTTCGAGAGCTGCTGCAGACCAAGACGCGCGGCGCTGCTGGAGGG 54035
QY 61 GCTGAGCGTGGGAGCTGCTGCGGAGAGCGTCCCTTCATCAGACCATCGGCT 120
Db GCTGAGCGTGGGAGCTGCTGCGGAGAGCGTCCCTTCATCAGACCATCGGCT 120
QY 54036 GCTGAGCGTGGGAGCTGCTGCGGAGAGCGTCCCTTCATCAGACCATCGGCT 54095
Db GCTGAGCGTGGGAGCTGCTGCGGAGAGCGTCCCTTCATCAGACCATCGGCT 54095
QY 121 GCTGCGGCCAGTGCCTCGCCCGGAGCGGAGCGGCGCTGAGCGGCGGCT 180
Db GCTGCGGCCAGTGCCTCGCCCGGAGCGGAGCGGCGCTGAGCGGCGGCT 180
QY 54096 GCTGCGGCCAGTGCCTCGCCCGGAGCGGAGCGGCGCTGAGCGGCGGCT 54155
Db GCTGCGGCCAGTGCCTCGCCCGGAGCGGAGCGGCGCTGAGCGGCGGCT 54155
QY 181 GCCCGCGCGTCCCGAGGAGTGCCTTCGAGCGGAGTGCAGTACACGGGCGCTT 240
Db GCCCGCGCGTCCCGAGGAGTGCCTTCGAGCGGAGTGCAGTACACGGGCGCTT 240
QY 241 CCACCTGGCCATCGAGTGA 261
Db CCACCTGGCCATCGAGTGA 54236
RESULT 7
BD169701 3847 bp DNA linear PAT 17-JAN-2003
LOCUS Human glioma antigen and method of preparing the same.
DEFINITION BD169701
ACCESSION BD169701
VERSION BD169701.1 GI:27875513
KEYWORDS WO 02055695-A/6.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Toda, M., Kawakami, Y., Kawase, T. and Iizuka, Y.
JOURNAL Human glioma antigen and method of preparing the same
Patent: WO 02055695-A 6 18-JUL-2002;
KEIO UNIVERSITY, MASAHIRO TODA, YUTAKA KAWAKAMI, TAKESHI KAWASE,
YUKIHIKO IIZUKA
COMMENT OS Homo sapiens (human)
PN WO 02055695-A/6
PD 18-JUL-2002
PF 30-NOV-2001 WO 2001JP010505
PR 09-JAN-2001 JP QIP 001965
PI MASAHIRO TODA, YUTAKA KAWAKAMI, TAKESHI KAWASE, YUKIHIKO IIZUKA
PC C12N15/12, C12N5/10, A01K67/027, A61K31/711, A61K38/00, A61K39/00,
A61K39/395,
```

```
PC A61K45/00, A61K48/00, A61P35/00, C07K14/47, C07K16/18, C07K19/00,
PC C12Q1/02,
PC C12Q1/68
CC Human glioma antigen and method of preparing the same FH Key
FEATURES Location/Qualifiers
FT CDS (211) (2883).
source 1..3847
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 18.0%; Score 47; DB 6; Length 3847;
Best Local Similarity 100.0%; Pred. No. 1e-12;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 215 GCGGAGTGGAGTACACGGGGCTTCACCTGGCCATCGACGTGA 261
Db GCGGAGTGGAGTACACGGGGCTTCACCTGGCCATCGACGTGA 47
RESULT 8
AC108407/c
AC108407 68196 bp DNA linear HTG 27-JAN-2002
LOCUS Mus musculus clone RP24-422P10, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC108407
ACCESSION AC108407.1 GI:18377216
VERSION HTG: HTGS PHASR0.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Mus musculus, clone RP24-422P10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 68196)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepe, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R.,
Landers, T., Lehotzky, J., Levine, R., Liu, G., MacLean, C.,
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuppback, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (27-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L18881
```

Center clone name: 422_P_10

* NOTE: This record contains 85 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
* 715 714: contig of 714 bp in length
* 814: gap of 100 bp
* 815 1520: contig of 706 bp in length
* 1521 1620: gap of 100 bp
* 1621 2323: contig of 703 bp in length
* 2324 2423: gap of 100 bp
* 2424 3134: contig of 711 bp in length
* 3135 3234: gap of 100 bp
* 3235 3937: contig of 703 bp in length
* 3938 4037: gap of 100 bp
* 4038 4739: contig of 702 bp in length
* 4740 4839: gap of 100 bp
* 4840 5545: contig of 706 bp in length
* 5546 6352: contig of 707 bp in length
* 6353 6452: gap of 100 bp
* 6453 7172: contig of 720 bp in length
* 7173 7272: gap of 100 bp
* 7273 7596: contig of 724 bp in length
* 7597 8096: gap of 100 bp
* 8097 8814: contig of 718 bp in length
* 8815 9629: contig of 715 bp in length
* 9630 9729: gap of 100 bp
* 9730 10417: contig of 688 bp in length
* 10418 10517: gap of 100 bp
* 10518 11223: contig of 706 bp in length
* 11224 11323: gap of 100 bp
* 11324 12014: contig of 691 bp in length
* 12015 12114: gap of 100 bp
* 12115 12816: contig of 702 bp in length
* 12817 12917: gap of 100 bp
* 12918 13616: contig of 700 bp in length
* 13617 13716: gap of 100 bp
* 13717 14422: contig of 706 bp in length
* 14423 15234: contig of 712 bp in length
* 15235 15334: gap of 100 bp
* 15335 16043: contig of 709 bp in length
* 16044 16143: gap of 100 bp
* 16144 16859: contig of 716 bp in length
* 16860 16959: gap of 100 bp
* 16960 17669: contig of 710 bp in length
* 17670 17769: gap of 100 bp
* 17770 18460: contig of 691 bp in length
* 18461 18560: gap of 100 bp
* 18561 19269: contig of 709 bp in length
* 19270 19369: gap of 100 bp
* 19370 20070: contig of 701 bp in length
* 20071 20170: gap of 100 bp
* 20171 20874: contig of 704 bp in length
* 20875 20974: gap of 100 bp
* 20975 21684: contig of 710 bp in length
* 21685 21784: gap of 100 bp
* 21785 22491: contig of 707 bp in length
* 22492 22591: gap of 100 bp
* 22592 23299: contig of 708 bp in length
* 23300 23399: gap of 100 bp
* 23400 24086: contig of 687 bp in length
* 24087 24186: gap of 100 bp

24187 24901: contig of 715 bp in length
24902 25001: gap of 100 bp
25002 25714: contig of 713 bp in length
25715 25814: gap of 100 bp
25815 26528: contig of 714 bp in length
26529 27358: contig of 730 bp in length
26629 27458: gap of 100 bp
27359 28159: contig of 701 bp in length
27459 28259: gap of 100 bp
28160 28966: contig of 707 bp in length
28260 29066: gap of 100 bp
28967 29765: contig of 699 bp in length
29067 29865: gap of 100 bp
29766 29866 29865: contig of 712 bp in length
29866 30677: gap of 100 bp
30677 31381: contig of 704 bp in length
31381 31481: gap of 100 bp
31482 32192: contig of 711 bp in length
32193 32292: gap of 100 bp
32293 32974: contig of 682 bp in length
32975 33074: gap of 100 bp
33075 33760: contig of 686 bp in length
33761 33860: gap of 100 bp
33861 34574: contig of 714 bp in length
34575 34674: gap of 100 bp
34675 35378: contig of 704 bp in length
35379 36194: contig of 716 bp in length
36195 36294: gap of 100 bp
36295 36973: contig of 679 bp in length
36974 37073: gap of 100 bp
37074 37774: contig of 701 bp in length
37775 37874: gap of 100 bp
37875 38586: contig of 712 bp in length
38587 38686: gap of 100 bp
38687 39388: contig of 702 bp in length
39389 39488: gap of 100 bp
39489 40195: contig of 707 bp in length
40196 40295: gap of 100 bp
40296 41013: contig of 718 bp in length
41014 41113: gap of 100 bp
41114 41828: contig of 715 bp in length
41829 41928: gap of 100 bp
41929 42643: contig of 715 bp in length
42644 42743: gap of 100 bp
42744 43458: contig of 715 bp in length
43459 43558: gap of 100 bp
43559 44242: contig of 684 bp in length
44243 44342: gap of 100 bp
44343 45043: contig of 701 bp in length
45044 45143: gap of 100 bp
45144 45856: contig of 713 bp in length
45857 45956: gap of 100 bp
45957 46654: contig of 698 bp in length
46655 46754: gap of 100 bp
46755 47449: contig of 695 bp in length
47450 47549: gap of 100 bp
47550 48258: contig of 709 bp in length
48259 48358: gap of 100 bp
48359 49070: contig of 712 bp in length
49071 49170: gap of 100 bp
49171 49886: contig of 716 bp in length
49887 49986: gap of 100 bp
49987 50673: contig of 687 bp in length
50674 50773: gap of 100 bp
50774 51472: contig of 699 bp in length
51473 51572: gap of 100 bp
51573 52288: contig of 716 bp in length
52289 52388: gap of 100 bp
52389 53075: contig of 687 bp in length
53076 53175: gap of 100 bp
53176 53879: contig of 704 bp in length

TITLE
JOURNAL
REFERENCE
AUTHORS
 Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (31-MAY-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 4 (bases 1 to 221524)
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Brauch, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (16-JUL-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 16, 2003 this sequence version replaced gi:31249817.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L29063
 Center clone name: 292_H_20

FEATURES
 source
 Location/Qualifiers
 1..221524
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="19"
 /map="19"
 /clone="RP23-292H20"
 /clone_lib="RPC1-23 Female Mouse BAC"
 misc_feature
 1..7046
 /note="wgs_end_extension
 clone_end:SP6"
 1234..1239
 /note="x30 qual SINGL region"
 repeat_region
 complement(1312..1743)
 /rpt_family="MLT1H1"
 repeat_region
 complement(1981..2123)
 /rpt_family="MER117"
 2143..2511
 /rpt_family="ORR1D"
 2897..2955
 /rpt_family="(CA)n"
 repeat_region
 3063..3114
 /rpt_family="(TG)n"
 repeat_region
 3141..3217
 /rpt_family="(CA)n"
 complement(3235..3402)
 /rpt_family="B4A"
 repeat_region
 complement(3447..3584)
 /rpt_family="B1 MM"
 repeat_region
 complement(3456..3690)

```
repeat_region /rpt_family="ID_B1"
repeat_region 3724..3752
repeat_region /rpt_family="(A)n"
repeat_region complement(4335..4351)
repeat_region /rpt_family="B3A"
repeat_region 4352..4408
repeat_region /rpt_family="(TG)n"
repeat_region complement(4409..4625)
repeat_region /rpt_family="B3A"
repeat_region complement(4701..4799)
repeat_region /rpt_family="MIR"
repeat_region complement(5000..5112)
repeat_region /rpt_family="B1F"
repeat_region 5268..5431
repeat_region /rpt_family="ORR1D"
repeat_region 5782..5870
repeat_region /rpt_family="B4"
repeat_region 5897..6040
repeat_region /rpt_family="B1_MM"
repeat_region 6447..6481
repeat_region /rpt_family="(TC)n"
repeat_region 6481..6516
repeat_region /rpt_family="(CA)n"
misc_feature 7047..7052
note="clone boundary
clone end:SP6
site:EcCK1"
repeat_region 7196..7246
repeat_region /rpt_family="(GAAA)n"
repeat_region complement(7364..7656)
repeat_region /rpt_family="ORR1C"
repeat_region 8336..8484
repeat_region /rpt_family="B1_MM"
repeat_region 8968..9042
repeat_region /rpt_family="L2"
repeat_region complement(9139..9332)
repeat_region /rpt_family="MIR"
repeat_region complement(9553..9742)
repeat_region /rpt_family="B2_MM2"
repeat_region complement(9847..10055)
repeat_region /rpt_family="B3"
repeat_region 11638..11844
repeat_region /rpt_family="B3"
repeat_region 12218..12494
repeat_region /rpt_family="RMER14"
repeat_region complement(13731..13993)
repeat_region /rpt_family="B4"
repeat_region 14922..14986
repeat_region /rpt_family="MIR"
repeat_region 15105..15218
repeat_region /rpt_family="LIME3"
repeat_region 15345..15509
repeat_region /rpt_family="LIMC4"
repeat_region complement(15704..15731)
repeat_region /rpt_family="ID_B1"
repeat_region 15732..15849
repeat_region /rpt_family="ORR1D"
repeat_region 15850..15874
repeat_region /rpt_family="(TGGGGG)n"
repeat_region 15875..16100
repeat_region /rpt_family="ORR1D"
repeat_region complement(16101..16251)
repeat_region /rpt_family="ID_B1"
repeat_region 17473..17513
repeat_region /rpt_family="(CA)n"
repeat_region 17626..17653
repeat_region /rpt_family="(CCCA)n"
repeat_region 18382..18451
repeat_region /rpt_family="ID4"
repeat_region complement(18592..18730)
repeat_region /rpt_family="RMER10A"

Best Local Similarity 100.0%; Pred. No. 7.5e-12;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGGTGAGTTCGAGGAGCTGCTGCAGACCAAGCGCGG 45
|||||
Db 42833 GATCAAGGTGAGTTCGAGGAGCTGCTGCAGACCAAGCGCGG 42877

RESULT 10
AC127769 243412 bp DNA linear HTG 19-NOV-2002
LOCUS Rattus norvegicus clone CH230-157C16, WORKING DRAFT SEQUENCE, 3
DEFINITION unordered pieces.
ACCESSION AC127769
VERSION AC127769.3 GI:25079525
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (Bases 1 to 243412)
Muzny,D.Marie., Metzker,M.Lee., Abranzen,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswato,K., Blair,J., Biankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagge,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Georgiadis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,M.,
Guaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Haviak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Louised,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munday,M., Murphy,M., Naif,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwackemehe,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poidexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villanueva,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,D., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,G., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von
Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct submission
TITLE
JOURNAL
Unpublished
```

```

REFERENCE
AUTHORS      2 (bases 1 to 243412)
TITLE        Worley,K.C.
JOURNAL      Direct Submission
             Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
REFERENCE    3 (bases 1 to 243412)
AUTHORS      Rat Genome Sequencing Consortium.
TITLE        Direct Submission
JOURNAL      Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
COMMENT      On Nov 19, 2002 this sequence version replaced gi:23269487.
             The sequence in this assembly is a combination of BAC based reads
             and whole genome shotgun sequencing reads assembled using Atlas
             (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
             in the feature table below represents a scaffold in the Atlas
             assembly (a 'contig-scaffold'). Within each contig-scaffold,
             individual sequence contigs are ordered and oriented, and separated
             by sized gaps filled with Ns to the estimated size. The sequence
             may extend beyond the ends of the clone and there may be sequence
             contigs within a contig-scaffold that consist entirely of whole
             genome shotgun sequence reads. Both end sequences and whole genome
             shotgun sequence only contigs will be indicated in the feature
             table.
             ----- Genome Center
             Center: Baylor College of Medicine
             Center code: BCM
             Web site: http://www.hgsc.bcm.tmc.edu/
             Contact: hgsc-help@bcm.tmc.edu
             ----- Project Information
             Center project name: GZSV
             Center clone name: CH230-157C16
             ----- Summary Statistics
             Assembly program: Phrap; version 0.990329
             Consensus quality: 220496 bases at least Q40
             Consensus quality: 222898 bases at least Q30
             Consensus quality: 224623 bases at least Q20
             Estimated insert size: 224491; sum-of-contigs estimation
             Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
             -----
             * NOTE: Estimated insert size may differ from sequence length
             * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 3 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence.
             * as soon as it is available and the accession number will
             * be preserved.
             *
             * 1 11919: contig of 11919 bp in length
             * 11920 12019: gap of unknown length
             * 12020 18301: contig of 6282 bp in length
             * 18302 18401: gap of unknown length
             * 18402 243412: contig of 225011 bp in length.
             FEATURES
             source
             Location/Qualifiers
               1. .243412
                 /organism="Rattus norvegicus"
                 /mol_type="genomic DNA"
                 /db_xref="taxon:10116"
                 /clone="CH230-157C16"
             misc_feature
             1. .1535
               /note="wgs contig"
             misc_feature
             12020. .13436
               /note="wgs contig"
             misc_feature
             17209. .18301
               /note="wgs contig"
             misc_feature
             18402. .19442
               /note="wgs contig"
             misc_feature
             25075. .27269
               /note="wgs contig"
             misc_feature
             238399. .242288
               /note="wgs contig"

```

```

misc_feature /note="wgs contig"
242339. .243412
/note="wgs contig"

ORIGIN
Query Match 17.2%; Score 45; DB 2; Length 243412;
Best Local Similarity 100.0%; Pred. No. 7.4e-12;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAACGACGCGG 45
    |||
Db 90830 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAACGACGCGG 90874

RESULT 11
BX511030 179734 bp DNA linear HTG 30-JUN-2003
LOCUS     BX511030.3 GI:32399530
DEFINITION
            Danio rerio clone CH211-66120, WORKING DRAFT SEQUENCE, 3 unordered
            pieces.
ACCESSION BX511030
VERSION   BX511030.3 GI:32399530
KEYWORDS  HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE    Danio rerio (zebrafish)
ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 179734)
AUTHORS   McLaren,S.
TITLE     Direct Submission
JOURNAL   Submitted (26-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfsh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
            On Jul 1, 2003 this sequence version replaced gi:31071341.
COMMENT   ----- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfsh-help@sanger.ac.uk
            ----- Project Information
            Center project name: zC66120
            ----- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 178170 bases at least Q40
            Consensus quality: 178389 bases at least Q30
            Consensus quality: 178746 bases at least Q20
            Insert size: 179534; sum-of-contigs
            Insert coverage: 9.1lx in Q20 bases; sum-of-contigs Quality
            coverage: 9.46x in Q20 bases; agarose-fp
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved
            *
            * 1 89039: contig of 89039 bp in length
            * 89040 89139: gap of 100 bp
            * 89140 169174: contig of 80035 bp in length
            * 169175 169274: gap of 100 bp
            * 169275 179734: contig of 10460 bp in length.
            FEATURES
            source
            Location/Qualifiers
               1. .179734
                 /organism="Danio rerio"
                 /mol_type="genomic DNA"
                 /db_xref="taxon:7955"
                 /clone="CH211-66120"
                 /clone_lib="CHORI-211"
             misc_feature 1. .89039

```

/note="assembly_fragment:02405
fragment chain:1
clone_end:17
vector_side:left"
89140..169174
/note="assembly_fragment:01570
fragment chain:1
clone_end:SP6
vector_side:right"
169275..179734
/note="assembly_fragment:00769"

misc_feature

misc_feature

ORIGIN

Query Match 8.8%; Score 23; DB 2; Length 179734;
Best Local Similarity 100.0%; Pred. No. 0.96; 0; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 0;

QY 41 GCCGGCGCCTGCTGGAGGGCT 63
|||||
DB 167280 GCCGGCGCCTGCTGGAGGGCT 167302

RESULT 12
AF076980
LOCUS AF076980 3291 bp DNA linear BCT 03-MAR-1999
DEFINITION Streptomyces albus G heat shock protein ClpB (clpB) gene, complete cds.
ACCESSION AF076980
VERSION AF076980.1 GI:4322268
KEYWORDS
SOURCE Streptomyces albus G
ORGANISM Streptomyces albus G
REFERENCE 1 (bases 1 to 3291)
AUTHORS Grandvalet,C., de Crecy-Lagard,V. and Mazodier,P.
TITLE The ClpB ATPase of Streptomyces albus G belongs to the HspR heat shock regulon
JOURNAL Mol. Microbiol. 31 (2), 521-532 (1999)
MEDLINE 99157559
PUBMED 10027969
REFERENCE 2 (bases 1 to 3291)
AUTHORS Grandvalet,C., de Crecy-Lagard,V. and Mazodier,P.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-1998) Biochimie Microbienne, Institut Pasteur, 25, rue du Docteur Roux, Paris 75724, France
FEATURES
source
1..3291
/organism="Streptomyces albus G"
/mol_type="genomic DNA"
/db_xref="taxon:1962"

-35_signal repeat_region
257..262
271..289
/note="HspR associated inverted repeat; binds HspR repressor"
/rpt_family="HAIR"
/rpt_type=inverted

-10_signal gene
281..286
457..3030
457..3030

CDS

/gene="clpB"
/gene="clpB"
/note="ATPase"
/codon_start=1
/transl_table=11
/product="heat shock protein ClpB"
/protein_id="AA015989.1"
/db_xref="GI:4322269"
/translation="MDAELTNRDALNAATRAVSAGNPDLTPAHLILLLEGODNE
NLVDLLAVERPGGGSAPASLSLPGVTGTVAPQPNRDLLAVTADAGRAKD
LGDEFLSTEHLIGIRPTAPRPCSPGRAPTEKLEAFQNTGRGRRVTPRPEGQYK
ALEKFGTDFAARERKLDVIGRQDEIRVVQVLSRFRKNNFVLIGRPGVKTAVE
GLAQRIVGDPVESLKDRLVSLDLGAMVAGAKYRGFEERLTKTVLSEIKESDQGLIVT

FIDELHTVVGAGAADSAMDNMLKPMIARGELRMVGATTLLDEYERIEKDPALERRF
QQVLVAPSVSDSIALLGLGRVEAHHKVOIADSAIVAAATLSDRYITSRFLPKAI
DLVDEAASRLMEIDSSPLEIDELQSVDRLEKMEALDRDTPASRORLEKLRDLA
DRERSCAHRPEWEKQSLNRVGBELKERLDEIRGQAEAAQOQGFDTASKLLYGELPT
LERDLRWPABEEAAKDTMVKEVGPDADIADVVGSWTGPAGRLLEGETOKLLRWEAE
LGRRLIGQSEAVQAVSDAVRTRAGIADPDRTGSEFLGPTGVGKTELAKALADEFLD
DERAMIRIDMSEYGEKHSVARLVGAPGVGYEGGQLTEAVRRRPPYSVYLIDEVEKA
HPGVFDILLQVLDDBRLTDGQRTVDFRNTILVLTNSLGSOVLVGSAPPEEKRRQVME
VVRSEKPEFLNRDLDDLVIFSLADEDELARLAGIQIAGLARRLDRRLSLDVTPEALA
WIAKEGFDPAIGARPLRLIQTIAIGDRILAKELAGEVRDGTVRVDRVEDGLLVGRAE
G"

ORIGIN

Query Match 8.4%; Score 22; DB 1; Length 3291;
Best Local Similarity 100.0%; Pred. No. 4.2; 0; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;

QY 39 CGCGCGCGCGCTGCTGGAGGG 60
|||||
DB 2099 CGCGCGCGCGCTGCTGGAGGG 2120

RESULT 13
AR063085
LOCUS AR063085 1425 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 6 from patent US 5844079.
ACCESSION AR063085
VERSION AR063085.1 GI:5990776
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1425)
AUTHORS Ingham,P.W., McMahon,A.P. and Tabin,C.J.
TITLE Vertebrate embryonic pattern-inducing proteins, and uses related thereto
JOURNAL Patent: US 5844079-A 6 01-DEC-1998;
FEATURES
source
1..1425
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 8.0%; Score 21; DB 6; Length 1425;
Best Local Similarity 100.0%; Pred. No. 14; 0; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 0;

QY 139 CTCGCCACCGCGGAGCCCGA 159
|||||
DB 837 CTCGCCACCGCGGAGCCCGA 857

RESULT 14
AR122612
LOCUS AR122612 1425 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 6 from patent US 6165747.
ACCESSION AR122612
VERSION AR122612.1 GI:14106929
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1425)
AUTHORS Ingham,P.W., McMahon,A.P., Tabin,C.J., Bumcrot,D.A. and Marti-Gorostiza,E.
TITLE Nucleic acids encoding hedgehog proteins
JOURNAL Patent: US 6165747-A 6 26-DEC-2000;
FEATURES
source
1..1425
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 8.0%; Score 21; DB 6; Length 1425;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 CTCGGCCACCGGGAGCCCGA 159
|||||
Db 837 CTCGGCCACCGGGAGCCCGA 857

RESULT 15
ARI64240
LOCUS ARI64240 1425 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 6 from patent US 6271363.
ACCESSION ARI64240
VERSION ARI64240.1 GI:16235295
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1425)
AUTHORS Ingham, P.W., McMahon, A.P. and Tabin, C.J.
TITLE Nucleic acids encoding hedgehog proteins
JOURNAL Patent: US 6271363-A 6 07-AUG-2001;
FEATURES
source 1..1425
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 8.0%; Score 21; DB 6; Length 1425;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 CTCGGCCACCGGGAGCCCGA 159
|||||
Db 837 CTCGGCCACCGGGAGCCCGA 857

Search completed: July 18, 2004, 15:18:48
Job time : 1615 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 11:40:25 ; Search time 208.026 Seconds
(without alignments)
4583.781 Million cell updates/sec

Title: US-10-051-769-5

Perfect score: 22

Sequence: 1 aaggtagagtcgaggagctgc 22

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rod.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	22	100.0	2623	9	BC012186	BC012186 Homo sapi
2	22	100.0	2639	6	AX747159	AX747159 Sequence
3	22	100.0	2639	9	AK091746	AK091746 Homo sapi
4	22	100.0	3856	9	BC028375	BC028375 Homo sapi
5	22	100.0	49052	9	AL359836	AL359836 Human DNA
6	22	100.0	68196	2	AC108407	AC108407 Mus muscu
7	22	100.0	120578	9	AC005887	AC005887 c1b_173
8	22	100.0	221524	10	AC139040	AC139040 Mus muscu
9	22	100.0	243412	2	AC127769	AC127769 Rattus no
10	19.4	88.2	139517	9	AC105446	AC105446 Homo sapi
11	19	86.4	295150	1	SC0939126	AL939126 Streptomy
12	18.8	85.5	274	3	AF110372	AF110372 Leishmani
13	18.8	85.5	704	3	LEIGP9692A	M63109 Leishmania
14	18.8	85.5	2315	3	AY051720	AY051720 Drosophil
15	18.8	85.5	2905	3	AY059441	AY059441 Drosophil
16	18.8	85.5	4394	1	HPU80806	U80806 Hydrogenoph
17	18.8	85.5	4741	1	AF489516	AF489516 Methyloba
18	18.8	85.5	5283	9	AB040942	AB040942 Homo sapi
19	18.8	85.5	6245	3	AY198410	AY198410 Leishmani
20	18.8	85.5	18666	5	AF469049	AF469049 Gallus ga
21	18.8	85.5	60514	2	AC014121	AC014121 Drosophil
22	18.8	85.5	76448	2	AC026266	AC026266 Homo sapi
23	18.8	85.5	79977	2	AC012693	AC012693 Drosophil
24	18.8	85.5	110000	2	LMF1CHR32_02	Continuation (3 of
25	18.8	85.5	159413	2	AC018520	AC018520 Homo sapi
26	18.8	85.5	160325	9	AC097347	AC097347 Homo sapi
27	18.8	85.5	169457	2	AC009216	AC009216 Drosophil
28	18.8	85.5	169856	3	AC012162	AC012162 Drosophil
29	18.8	85.5	170000	2	AC004394	AC004394 Homo sapi
30	18.8	85.5	177835	3	AC091127	AC091127 Drosophil
31	18.8	85.5	192399	3	AC010032	AC010032 Drosophil
32	18.8	85.5	203777	9	CNS01DVH	AL135818 Human chr
33	18.8	85.5	207018	2	AC116539	AC116539 Drosophil
34	18.8	85.5	210515	9	AC097382	AC097382 Homo sapi
35	18.8	85.5	299804	3	AE003507	AE003507 Drosophil
36	18.8	85.5	318699	3	AB003470	AB003470 Drosophil
37	18.4	83.6	959	8	AB007506	AB007506 Triticum
38	18.4	83.6	2858	3	AY058563	AY058563 Drosophil
39	18.4	83.6	110032	2	AC014916	AC014916 Drosophil
40	18.4	83.6	132741	9	AL158043	AL158043 Human DNA
41	18.4	83.6	159672	3	AC010040	AC010040 Drosophil
42	18.4	83.6	163805	9	AC010956	AC010956 Homo sapi
43	18.4	83.6	166935	2	AC125957	AC125957 Rattus no
44	18.4	83.6	185956	2	AC129582	AC129582 Mus muscu
45	18.4	83.6	191494	10	AL596182	AL596182 Mouse DNA

ALIGNMENTS

RESULT 1
BC012186
LOCUS BC012186 2623 bp mRNA linear PRI 06-AUG-2001
DEFINITION Homo sapiens, clone IMAGE:4564853, mRNA.
ACCESSION BC012186
VERSION BC012186.1 GI:15082555
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2623)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapsb@mail.nih.gov
 Tissue Procurement: DCTD/DTF
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Genome Sequence Centre,
 BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
 Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
 Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
 Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
 Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
 Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
 George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 29 Row: h Column: 17
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis, Genomescan gene prediction.

FEATURES

source

```
1..2623
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4564853"
/tissue_type="kidney, renal cell adenocarcinoma"
/clone_lib="NIH MGC 14"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
```

ORIGIN

Query Match 100.0%; Score 22; DB 9; Length 2623;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22

DB 547 AAGGTGGAGTTCGAGGAGCTGC 568

RESULT 2

AX747159

LOCUS

AX747159 2639 bp mRNA linear PAT 20-JUN-2003

DEFINITION Sequence 684 from Patent EP1308459.

ACCESSION AX747159

VERSION AX747159.1 GI:32131547

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota

Mammalia

Primates

Catarrhini

Hominidae

Homo

Euteleostomi

Mammalia

Hominidae

Homo

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 2639;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22

DB 583 AAGGTGGAGTTCGAGGAGCTGC 604

RESULT 3

AX091746

LOCUS

AX091746

DEFINITION

Homo sapiens cDNA FLJ34427 fis, clone HHDPC2008816.

ACCESSION

AX091746

VERSION

AK091746.1 GI:21750192

KEYWORDS

oligo capping; fis (full insert sequence).

SOURCE

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

```

BC028375.1 GI:22382223
MGC.
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L.D., Shenmen,C.M., Schuler,G.D.,
Ahteshul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diachenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schneitz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carminci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalhus,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
MEDLINE 12477932
PUBMED
REFERENCE
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (23-APR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-i@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 34 Row: m Column: 21
This clone was selected for full length sequencing because it
passed the following selection criteria: GenomesScan gene
prediction.
FEATURES
source Location/Qualifiers
1 .3856
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:27107 IMAGE:4837939"
/tissue_type="Testis"
/clone_lib="NIH MGC_97"
/lab_host="DH10B"
/note="vector: pBluescript"
1 .3856
/genes="LOC118987"
/db_xref="LocusID:118987"
203 .3667
/codon_start=1
/product="hypothetical protein LOC118987"
/protein_id="AAH28375.1"

```

```

/db_xref="GI:22382224"
/db_xref="LocusID:118987"
/translation="MGLLMLILASAVLGSFLTLAQFFLYRQRPPEPPADEARAAGRG
FYRIKVPGLLRRELYLGGRRDEEPAEGGATAAPETPAPTRCTCYFLNATI
LRLFRPDRATLRTWVTKIKVEFELLQTKTAGRLGLSLGLDVLFLAGETVFFIKTI
RLVFRVPVSGTEPDGPEALPAACEPELAFAEAEVYNGSFLHLDVLFVKGSAYL
FVKLSRVGLRLVTRVPFTWHFFSFVEDPLIDREVRQFQEGRPMLQTSIIYNOLK
KIKRKHTLPYKIRKFPFPYQTLQGFPEDEHHIQWALTEGRLKVLLECSRLL
IFGSDREANVHCTLELSSVWEEKORSIKITVELIKGNLQSVGLRLVLYSDGAG
HIIETAVNSPAAATLQGRDLIAIGVKITSTLVKLIKQAGDRVLVYDVGAG
QNGQIAQVQNFQLEENFSSSCQSGYEAAAGLVDTSESLDEFFEDLADVRAQ
NEFKBOAQLSHSKRPVPTTSLKPLGASVPLNRKLVAGSHPLPKIOSKQKGVPP
LFTSEITDAQVSKPTQGSAPKPPVPRQAKVPLPSADAPNOAEPDVLEPEKVP
PELVDSASKQAKNVDAIDAAAPKQFLAKQEVAKVTSETSCPTKDSDDRTWESS
EILYRNKLGKTRTRASCLFIDIEACHYRLINLALWCRDPFKGLGICLGHVSLKLEDA
LGLCATNTEYLSKRLLEAPSKAIVTRALNLSMKGKFNKFCYDITIHFKYLKE
GESDHHVTVNVEKEPEHLVEEVSVLPKEEQFVGQGLTENKHSFQDITQNPWTCDY
CKKKWTKAASOCMFCAYVCHKKQCKLAETSVCAGTDRRIDRTLNLRLEQETLL
GLPPRVDAEASKVNKTGLTSHINTSSRLNLROVSKTRLSEPGTDILVEPSKHTP
NSTDNGSDEVCGNPSKRGNSGILKVRKEGLDDSVFIAVKEIGRLDLYRGLPTE
ERIKLEFMDLKLQNEIDQELHNLSVREEKETTDTTRKSLLSAALAKSGERLQALT
LMIHVRAGIEDIETLESLLQHSKKISKITYDDTEEDLDNEISLQDSQFSSISDD
LFGPSSSV"
1292 .1546
/misc_feature
/Note="PDZ; Region: Domain present in PSD-95, Dlg, and
ZO-1/2. Also called DHR (Dlg homologous region) or GLGF
(relatively well conserved tetrapeptide in these domains).
Some PDZs have been shown to bind C-terminal polypeptides".
/db_xref="CDD:smart00228"
2723 .2851
/misc_feature
/Note="DAG PE-bind; Region: Phorbol esters/diacylglycerol
binding domain (C1 domain). This domain is also known as
the Protein Kinase C conserved region 1 (C1) domain"
/db_xref="CDD:pfam00130"
ORIGIN
Query Match 100.0%; Score 22; DB 9; Length 3856;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGGTGAGTTCGAGGAGCTGC 22
|||||
DB 572 AAGGTGAGTTCGAGGAGCTGC 593
|||||
RESULT 5
AL359836/c
LOCUS Human DNA sequence from clone RP11-38956 on chromosome 10, complete
sequence.
DEFINITION AL359836
ACCESSION AL359836
VERSION AL359836.16 GI:17977720
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Smith, M.
Direct Submission
Submitted (21-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
On Dec 23, 2001 this sequence version replaced gi:17384082.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=

```

30): an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr10> RP11-389E6 is from the library RP11-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBAC3.6 IMPORTANT: This sequence is not the entire insert of clone RP11-389E6 It may be shorter because we sequence overlapping sections only once, except for a short overlap. The true left end of clone CTA-109P11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

FEATURES

source

```

1. Location/Qualifiers
   1. 49052
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosome="10"
      /clone="RP11-389E6"
      /clone_lib="RP11-11.2"

```

ORIGIN

```

Query Match      100.0%; Score 22; DB 9; Length 49052;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTGGAGTTCGAGGAGCTGC 22
Db 29763 AAGGTGGAGTTCGAGGAGCTGC 29742

RESULT 6
AC108407/c
LOCUS AC108407 58136 bp DNA linear HTG 27-JAN-2002
DEFINITION Mus musculus clone RP24-422P10; LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC108407
VERSION AC108407.1 GI:18377216
KEYWORDS HTG; HTGS PHASE0.
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
1 (bases 1 to 68196)
Mus musculus, clone RP24-422P10
Unpublished
2 (bases 1 to 68196)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Batra,N., Bastien,V., Boguslavskiy,I., Bouckghalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczyk,J., Levine,R., Liu,G., Maclean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McSwan,P., McKernan,K., Meldrim,J., Menseu,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,K.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,

```

Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
 Robertt,M., Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S.,
 Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (27-JAN-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: Li8881
 Center clone name: 422_P10

 * NOTE: This record contains 85 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.
 * 1 714: contig of 714 bp in length
 * 715 814: gap of 100 bp
 * 815 1520: contig of 706 bp in length
 * 1521 1620: gap of 100 bp
 * 1621 2323: contig of 703 bp in length
 * 2324 2423: gap of 100 bp
 * 2424 3134: contig of 711 bp in length
 * 3135 3234: gap of 100 bp
 * 3235 3937: contig of 703 bp in length
 * 3938 4037: gap of 100 bp
 * 4038 4739: contig of 702 bp in length
 * 4740 4839: gap of 100 bp
 * 4840 5545: contig of 706 bp in length
 * 5546 5645: gap of 100 bp
 * 5646 6352: contig of 707 bp in length
 * 6353 6452: gap of 100 bp
 * 6453 7172: contig of 720 bp in length
 * 7173 7272: gap of 100 bp
 * 7273 7996: contig of 724 bp in length
 * 7997 8096: gap of 100 bp
 * 8097 8814: contig of 718 bp in length
 * 8815 8914: gap of 100 bp
 * 8915 9629: contig of 715 bp in length
 * 9630 9729: gap of 100 bp
 * 9730 10417: contig of 688 bp in length
 * 10418 10517: gap of 100 bp
 * 10518 11223: contig of 706 bp in length
 * 11224 11323: gap of 100 bp
 * 11324 12014: contig of 691 bp in length
 * 12015 12114: gap of 100 bp
 * 12115 12816: contig of 702 bp in length
 * 12817 12916: gap of 100 bp
 * 12917 13616: contig of 700 bp in length
 * 13617 13716: gap of 100 bp
 * 13717 14422: contig of 706 bp in length
 * 14423 14522: gap of 100 bp
 * 14523 15234: contig of 712 bp in length
 * 15235 15334: gap of 100 bp
 * 15335 16043: contig of 709 bp in length
 * 16044 16143: gap of 100 bp

TITLE
 JOURNAL
 COMMENT

* 16144 16859: contig of 716 bp in length
* 16860 16959: gap of 100 bp
* 16960 17669: contig of 710 bp in length
* 17670 17769: gap of 100 bp
* 17770 18460: contig of 691 bp in length
* 18461 18560: gap of 100 bp
* 18561 19269: contig of 709 bp in length
* 19270 19369: gap of 100 bp
* 19370 20070: contig of 701 bp in length
* 20071 20170: gap of 100 bp
* 20171 20874: contig of 704 bp in length
* 20875 20974: gap of 100 bp
* 20975 21684: contig of 710 bp in length
* 21685 21784: gap of 100 bp
* 21785 22491: contig of 707 bp in length
* 22492 22591: gap of 100 bp
* 22592 23299: contig of 708 bp in length
* 23300 23399: gap of 100 bp
* 23400 24086: contig of 687 bp in length
* 24087 24186: gap of 100 bp
* 24187 24901: contig of 715 bp in length
* 24902 25001: gap of 100 bp
* 25002 25714: contig of 713 bp in length
* 25715 25814: gap of 100 bp
* 25815 26528: contig of 714 bp in length
* 26529 26628: gap of 100 bp
* 26629 27358: contig of 730 bp in length
* 27359 27458: gap of 100 bp
* 27459 28159: contig of 701 bp in length
* 28160 28259: gap of 100 bp
* 28260 28966: contig of 707 bp in length
* 28967 29066: gap of 100 bp
* 29067 29765: contig of 699 bp in length
* 29766 29865: gap of 100 bp
* 29866 30577: contig of 712 bp in length
* 30578 30677: gap of 100 bp
* 30678 31381: contig of 704 bp in length
* 31382 31481: gap of 100 bp
* 31482 32192: contig of 711 bp in length
* 32193 32292: gap of 100 bp
* 32293 32974: contig of 682 bp in length
* 32975 33074: gap of 100 bp
* 33075 33760: contig of 686 bp in length
* 33761 33860: gap of 100 bp
* 33861 34574: contig of 714 bp in length
* 34575 34674: gap of 100 bp
* 34675 35378: contig of 704 bp in length
* 35379 35478: gap of 100 bp
* 35479 36194: contig of 716 bp in length
* 36195 36294: gap of 100 bp
* 36295 36973: contig of 679 bp in length
* 36974 37073: gap of 100 bp
* 37074 37774: contig of 701 bp in length
* 37775 37874: gap of 100 bp
* 37875 38586: contig of 712 bp in length
* 38587 38686: gap of 100 bp
* 38687 39388: contig of 702 bp in length
* 39389 39488: gap of 100 bp
* 39489 40195: contig of 707 bp in length
* 40196 40295: gap of 100 bp
* 40296 41013: contig of 718 bp in length
* 41014 41114: gap of 100 bp
* 41115 41828: contig of 715 bp in length
* 41829 41928: gap of 100 bp
* 41929 42643: contig of 715 bp in length
* 42644 42743: gap of 100 bp
* 42744 43458: contig of 715 bp in length
* 43459 43558: gap of 100 bp
* 43559 44242: contig of 684 bp in length
* 44243 44342: gap of 100 bp
* 44343 45043: contig of 701 bp in length
* 45044 45143: gap of 100 bp
* 45144 45856: contig of 713 bp in length

* 45857 45956: gap of 100 bp
* 45957 46854: contig of 698 bp in length
* 46855 46954: gap of 100 bp
* 46955 47449: contig of 695 bp in length
* 47450 47549: gap of 100 bp
* 47550 48258: contig of 709 bp in length
* 48259 48358: gap of 100 bp
* 48359 49070: contig of 712 bp in length
* 49071 49170: gap of 100 bp
* 49171 49886: contig of 716 bp in length
* 49887 49986: gap of 100 bp
* 49987 50673: contig of 687 bp in length
* 50674 50773: gap of 100 bp
* 50774 51472: contig of 699 bp in length
* 51473 51572: gap of 100 bp
* 51573 52288: contig of 716 bp in length
* 52289 52388: gap of 100 bp
* 52389 53075: contig of 687 bp in length
* 53076 53175: gap of 100 bp
* 53176 53879: contig of 704 bp in length
* 53880 53979: gap of 100 bp
* 53980 54675: contig of 696 bp in length
* 54676 54775: gap of 100 bp
* 54776 55477: contig of 702 bp in length
* 55478 55577: gap of 100 bp

Query Match 100.0%; Score 22; DB 2; Length 68196;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGTGGAGTTCGAGGAGCTGC 22

Db 17501 AAGTGGAGTTCGAGGAGCTGC 17480

RESULT 7

AC005887 120578 bp DNA linear PRI 05-NOV-1999
LOCUS Citb_173_i_12, complete sequence.

DEFINITION

AC005887

ACCESSION

AC005887.3 GI:6249675

VERSION

HTG

KEYWORDS

HTG

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 120578)

AUTHORS

Smith,D.R.

TITLE

Sequencing of Human Chromosome 10

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 120578)

AUTHORS

Smith,D.R.

TITLE

Direct Submission

JOURNAL

Submitted (11-DEC-1998)

Genome Therapeutics Corporation, 100 Beaver

Street, Waltham, MA 02154, USA

REFERENCE

3 (bases 1 to 120578)

AUTHORS

Smith,D.R.

TITLE

Direct Submission

Submitted (05-NOV-1999)

Genome Therapeutics Corporation, 100 Beaver

Street, Waltham, MA 02154, USA

Vector Sequence Clipped

On Nov 5, 1999 this sequence version replaced gi:4314331.

Location/Qualifiers

```

source
1. 120578
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/map="10q25"
/clone_lib="CIT987SK-1173112"

ORIGIN
Query Match 100.0%; Score 22; DB 9; Length 120578;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGTGGAGGATTCGAGGAGCTGC 22
|||||
Db 53980 AAGTGGAGGATTCGAGGAGCTGC 54001
|||||

RESULT 8
AC139040
LOCUS AC139040 221524 bp DNA linear ROD 16-JUL-2003
DEFINITION Mus musculus chromosome 19, clone RP23-292H20, complete sequence.
ACCESSION AC139040
VERSION AC139040.7 GI:32813610
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus chromosome 19, clone RP23-292H20
JOURNAL Unpublished
COMMENT 2 (bases 1 to 221524)
1 (bases 1 to 221524)

FEATURES
source
1. 221524
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="19"
/map="19"
/clone_lib="RP23-292H20"
/notes="wgs end extension"
1. 7016
clone_end:SP6-
1234..1239
/notes="<30 qual SNGL region"
complement(1312..1743)
/rpt_family="MLT1H1"
repeat_region
complement(1981..2123)
/rpt_family="MER117"
repeat_region
2143..2511
/rpt_family="ORR1D"
repeat_region
2897..2955
/rpt_family="(CA)n"
repeat_region
3063..3114
/rpt_family="(TG)n"
repeat_region
3141..3217
/rpt_family="(CA)n"
repeat_region
complement(3235..3402)
/rpt_family="B4A"

TITLE
JOURNAL
REFERENCE
AUTHORS
Submitted (23-JAN-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 221524)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collamore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (31-MAY-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 221524)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collamore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-JUL-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 16, 2003 this sequence version replaced gi:31249817.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Smit, A.F.A. & Green, P. (1996-1997)
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIMR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L29063
Center clone name: 292_H_20
----- Location/Qualifiers
1. 221524
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="19"
/map="19"
/clone_lib="RP23-292H20"
/notes="wgs end extension"
1. 7016
clone_end:SP6-
1234..1239
/notes="<30 qual SNGL region"
complement(1312..1743)
/rpt_family="MLT1H1"
repeat_region
complement(1981..2123)
/rpt_family="MER117"
repeat_region
2143..2511
/rpt_family="ORR1D"
repeat_region
2897..2955
/rpt_family="(CA)n"
repeat_region
3063..3114
/rpt_family="(TG)n"
repeat_region
3141..3217
/rpt_family="(CA)n"
repeat_region
complement(3235..3402)
/rpt_family="B4A"

```



```

repeat_region complement(3447..3584)
repeat_region /rpt_family="B1_MM"
repeat_region complement(3456..3690)
repeat_region /rpt_family="ID_B1"
repeat_region 3724..3752
repeat_region /rpt_family="(A)n"
repeat_region complement(4335..4351)
repeat_region /rpt_family="B3A"
repeat_region 4352..4408
repeat_region /rpt_family="(TG)n"
repeat_region complement(4409..4625)
repeat_region /rpt_family="B3A"
repeat_region complement(4701..4799)
repeat_region /rpt_family="MIR"
repeat_region complement(5000..5112)
repeat_region /rpt_family="B1F"
repeat_region 5268..5431
repeat_region /rpt_family="ORR1D"
repeat_region 5782..5870
repeat_region /rpt_family="B4"
repeat_region 5897..6040
repeat_region /rpt_family="B1_MM"
repeat_region 6447..6481
repeat_region /rpt_family="(TC)n"
repeat_region 6481..6516
repeat_region /rpt_family="(CA)n"
misc_feature 7047..7052
/notes="clone_boundary
clone_end:SP6
site:ECORI"
7196..7246
/rpt_family="(GAAA)n"
repeat_region complement(7364..7656)
repeat_region /rpt_family="ORR1C"
repeat_region 8336..8484
repeat_region /rpt_family="B1_MM"
repeat_region 8968..9042
/rpt_family="L2"
repeat_region complement(9139..9332)
repeat_region /rpt_family="MIR"
repeat_region complement(9553..9742)
repeat_region /rpt_family="B2_Mm2"
repeat_region complement(9847..10055)
repeat_region /rpt_family="B3"
repeat_region 11638..11844
/rpt_family="B3"
repeat_region 12218..12494
/rpt_family="RMER14"
repeat_region complement(13731..13993)
repeat_region /rpt_family="B4"
repeat_region 14922..14986
/rpt_family="MIR"
repeat_region 15105..15218
/rpt_family="L1ME3"
repeat_region 15345..15509
/rpt_family="L1MC4"
repeat_region complement(15704..15731)
repeat_region /rpt_family="ID_B1"
repeat_region 15732..15849
/rpt_family="ORR1D"
repeat_region complement(16101..16251)
repeat_region /rpt_family="ID_B1"
repeat_region 17479..17513
/rpt_family="(CA)n"
repeat_region 17626..17653
/rpt_family="(CCA)n"
repeat_region 18382..18451
/rpt_family="ID4"
repeat_region complement(18592..18730)

```

```

/rpt_family="RMER10A"

Query Match      100.0%; Score 22; DB 10; Length 221524;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGTGGAGTTCGAGGAGCTGC 22
    |||||
Db 42837 AAGGTGGAGTTCGAGGAGCTGC 42858

RESULT 9
AC127769
LOCUS      243412 bp DNA linear HTG 19-NOV-2002
DEFINITION Rattus norvegicus clone CH230-157C16, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
AC127769
AC127769.3 GI:25079525
VERSION    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS   Rattus norvegicus
SOURCE     Rattus norvegicus
ORGANISM   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 243412)
AUTHORS   Muzny,D,Marie., Metzker,M, Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregregis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,I., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokeme,O., Okwuonu,G., Olarnpungsoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,R., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6.

FEATURES

source

```
1. .139517
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /chromosome="7"
  /map="7"
```

```
/clone="RP11-395B7"
```

```
/clone_lib="RPCI-11"
```

```
290. .591
```

```
/rpt_family="Alu"
```

```
1071. .1105
```

```
/rpt_family="(GA)n"
```

```
1159. .1274
```

```
/rpt_family="Alu"
```

```
1275. .1312
```

```
/rpt_family="(TAA)n"
```

```
1313. .1471
```

```
/rpt_family="Alu"
```

```
2694. .3750
```

```
/note="CpG island (%GC=69.7, o/e=0.77, #CpGs=89)"
```

```
3678. .3700
```

```
/rpt_family="GC-rich"
```

```
3917. .4005
```

```
/rpt_family="CT-rich"
```

```
4453. .4587
```

```
/rpt_family="MIR"
```

```
5342. .5438
```

```
/rpt_family="C-rich"
```

```
7443. .7827
```

```
/rpt_family="Alu"
```

```
7843. .7973
```

```
/rpt_family="Alu"
```

```
7974. .8244
```

```
/rpt_family="Alu"
```

```
8245. .8418
```

```
/rpt_family="Alu"
```

```
9438. .9501
```

```
/rpt_family="MIR"
```

```
repeat_region
  /rpt_family="Alu"
  9748. .10043
repeat_region
  /rpt_family="MaLR"
  10702. .10781
repeat_region
  /rpt_family="Alu"
  10782. .10918
repeat_region
  /rpt_family="MaLR"
  10919. .10983
repeat_region
  /rpt_family="MaLR"
  11012. .11106
repeat_region
  /rpt_family="Alu"
  11110. .11393
repeat_region
  /rpt_family="Alu"
  11445. .11775
repeat_region
  /rpt_family="L1"
  11776. .12103
repeat_region
  /rpt_family="Alu"
  12104. .12150
repeat_region
  /rpt_family="L1"
  13244. .13530
repeat_region
  /rpt_family="Alu"
  13562. .13853
repeat_region
  /rpt_family="Alu"
  13858. .14010
repeat_region
  /rpt_family="MIR"
  14011. .14318
repeat_region
  /rpt_family="Alu"
  14319. .14417
repeat_region
  /rpt_family="MIR"
  14434. .14608
repeat_region
  /rpt_family="MaLR"
  14609. .14901
repeat_region
  /rpt_family="Alu"
  14959. .15105
repeat_region
  /rpt_family="Alu"
  15132. .15377
repeat_region
  /rpt_family="MaLR"
  15654. .15768
repeat_region
  /rpt_family="MIR"
  16056. .16335
repeat_region
  /rpt_family="Alu"
  16668. .16730
repeat_region
  /rpt_family="(TAAAAA)n"
  16732. .16737
repeat_region
  /rpt_family="(TA)n"
  16738. .17039
repeat_region
  /rpt_family="Alu"
  17040. .17089
repeat_region
  /rpt_family="(TA)n"
  17112. .17261
repeat_region
  /rpt_family="Alu"
  17269. .17579
repeat_region
  /rpt_family="Alu"
  17600. .17879
repeat_region
  /rpt_family="L1"
  17880. .18169
repeat_region
  /rpt_family="Alu"
  18170. .18417
repeat_region
  /rpt_family="L1"
  19105. .19404
repeat_region
  /rpt_family="Alu"
  19663. .19966
repeat_region
  /rpt_family="Alu"
  19967. .20272
repeat_region
  /rpt_family="Alu"
  20307. .20375
repeat_region
  /rpt_family="MIR"
  20376. .20675
repeat_region
  /rpt_family="Alu"
  20676. .20706
repeat_region
  /rpt_family="MIR"
  20814. .20889
repeat_region
  /rpt_family="L2"
  21151. .21234
```

Query Match 88.2%; Score 19.4; DB 9; Length 139517;
 Best Local Similarity 95.2%; Pred. No. 3.8e-02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAGTGGAGTTCAGGAGCTG 21
 |||||
 DB 104373 AAGTGGAGTTCAGGAGCTG 104393

RESULT 11
 LOCUS SC0939126 295150 bp DNA linear BCT 11-FEB-2003
 DEFINITION Streptomyces coelicolor A3(2) complete genome; segment 23/29.
 ACCESSION AL039126 AL009199 AL023517 AL023702 AL031514 AL034446 AL034492
 AL035212 AL035478 AL049727 AL132824 AL132997 AL589708 AL591522
 AL645882
 VERSION AL039126.1 GI:24413894

KEYWORDS Streptomyces coelicolor A3(2)
 SOURCE Streptomyces coelicolor A3(2)
 ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1
 AUTHORS Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L.,
 Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H.,
 Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M.,
 Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S.,
 Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neill, S.,
 Rabinowitz, E., Rajandream, M.A., Rutherford, K., Rutter, S.,
 Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S.,
 Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G.,
 Parkhill, J., and Hopwood, D.A.
 Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2)
 Nature 417 (6885), 141-147 (2002)

JOURNAL MEDLINE
 PUBMED 1200953
 REFERENCE 2 (bases 1 to 295150)
 AUTHORS Bentley, S.D.
 TITLE Direct Submission
 JOURNAL Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
 sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
 Hinxton, Cambridge CB10 1SA E-mail: sds@sanger.ac.uk
 On or before Oct 26, 2002 this sequence version replaced
 gi:20520744, gi:20520662, gi:20520663, gi:20520667, gi:20520669,
 gi:20520761, gi:20520816, gi:20520670, gi:20520820, gi:20520846,
 gi:20520823, gi:20520863, gi:20520869.

FEATURES
 source
 1..295150
 /organism="Streptomyces coelicolor A3(2)"
 /mol_type="genomic DNA"
 /strain="A3(2)"
 /db_xref="taxon:100226"
 512..515
 /note="possible RBS upstream of SC7B7.02"
 524..1933
 /gene="SCO6005"
 /note="synonym: SC7B7.02"
 524..1933
 /gene="SCO6005"
 /note="SC7B7.02, possible lipoprotein, len: 469; has
 similarity to signal sequence of Mycobacterium leprae
 hypothetical protein TR:E334742 (EMBL:Z98604) MLCB2052.27
 (445 aa), fasta scores; opt: 75 z-score: 199.2 E(1):
 0.00062, 23.3% identity in 335 aa overlap. Contains
 N-terminal signal sequence and appropriately positioned
 PS00013 Prokaryotic membrane lipoprotein lipid attachment
 site"
 /codon_start=1
 /transl_table=11
 /product="putative lipoprotein"
 /protein_id="CAAL5782.1"

/db_xref="GI:2661677"
 /db_xref="GOA:O50499"
 /db_xref="SPTREMBL:O50499"
 /translation="MTIRAGSLDRRLRLGALTAATMGSAFAVACSSPSSDEKESDGGP
 KGEKSAANPFGAANSTVEAAIFDGGYGTVDYVYANQVLGSGVKGKLVQVPPVVDIAP
 QLOPRFVGNPDIDNSGEDIQILGILDOLELDLDFEASTYEGKKIADIVPEGVK
 DGTGFKDFVALNVMTVYGVYKSTLFEENGWTPPKTWDRLDLGOEAKKKGLVLFV
 HGKEATYVRLLLIDSALEKGGDEVRLEALEKGCWSPHAPVQGVIKVMTWVKQKMF
 VPGSGTQFQKAQAIWSNDKALLYPSGWIENEMKKAATKADFQMTGTPSMILTDKPA
 LPYALRAAAGEPFIVPRQGNPAGKEVLRAMLSEKAAANFSLGAPITIVKGTVPA
 DYGSTALVSQTKMELEAAGTINIFNYFVETVGLNTDQLVPWNSFLAGLDGKGLTSL
 QKISDKVREDSDVDKVKVS"
 581..613
 /gene="SCO6005"
 /note="PS00013 Prokaryotic membrane lipoprotein lipid
 attachment site"
 1941..3002
 /gene="SCO6006"
 /note="synonym: SC7b7.03"
 1941..3002
 /gene="SCO6006"
 /note="SC7b7.03, probable integral membrane transport
 protein, len: 353 aa; similar to many members of the MALFG
 subfamily of the binding-protein-dependent transport
 system membrane proteins eg. MSMP STRMU Q00750 multiple
 sugar-binding transport system permease protein (290 aa),
 fasta scores; opt: 393 z-score: 636.8 E(1): 2.6e-28, 32.5%
 identity in 280 aa overlap. Contains PS00402
 Binding-protein-dependent transport systems inner membrane
 comp signature"
 /codon_start=1
 /transl_table=11
 /product="probable transmembrane transport protein"
 /protein_id="CAAL5783.1"
 /db_xref="GI:2661678"
 /db_xref="GOA:O50500"
 /db_xref="SPTREMBL:O50500"
 /translation="WKDTIPTAETASRRPEPAARGRRRKLTFDRVTFVFLAFLGVGP
 LAIPVFLVLPFGOALFMGMDTMRGSPDYNGFDNFTKMFQDDIFLKALRNALLA
 AFVLVTLTALGVAVAITLGGSPKPGVIRGASFYRLISFFPVVPAIVGLIWAQ
 MYDNAGLLNGVLTLGLDQDFTFALWKEKAAAPVAFVILGVGLFVAVIQAIAIK
 GVPGLYEAAKIDGAGRTTISITLPAIRDSVQTAYTILGIALDFAVYVQVAMPNG
 GPDNSTITISORLNFVAFKQQFGYATANGVLAATLVFAALVFLNRLTGSGEGES
 KRKAPGSRARRAAAKGGAR"
 2589..2675
 /gene="SCO6006"
 /note="PS00402 Binding-protein-dependent transport systems
 inner membrane component signature"
 2988..2992
 /gene="SCO6006"
 /note="possible RBS upstream of SC7b7.04"
 2999..3907
 /gene="SCO6007"
 /note="synonym: SC7b7.04"
 2999..3907
 /gene="SCO6007"
 /note="SC7b7.04, probable integral membrane transport
 protein, len: 301 aa; similar to many members of the MALFG
 subfamily of the binding-protein-dependent transport
 systems membrane proteins eg. TR:Q06111 (EMBL:X66092)
 putative transport system permease protein from
 Clostridium perfringens (275 aa), fasta scores; opt: 590
 z-score: 796.5 E(1): 0, 33.6% identity in 265 aa overlap.
 Contains PS00402 Binding-protein-dependent transport
 systems inner membrane comp signature"
 /codon_start=1
 /transl_table=11
 /product="probable transmembrane transport protein"
 /protein_id="CAAL5784.1"
 /db_xref="GI:2661679"
 /db_xref="GOA:O50501"
 /db_xref="SPTREMBL:O50501"
 /translation="MSVIAADKKPASDRKPVSDRKLARAAASSDRRPAISHALLILM

```
SVTVIVPMLWVLMSSFKSTCEILLSPFSLPDHWRPENYANAWTDANIGKFLNSVLIV
VSLIIVMLIGACAVLARFEPFGRRLIYYVMLAGLTFPVFLAIVPLFFQJONFGLL
NTRPGILTYAFALFTWFFLYFSPFSLPHDYLIDGAGDWEAFQVMDPMARP
GMAAVAFNFLGNQFLFLFVALNTDQDKWVLITQGMAYASSQVYDIDYGFALPAIVV
TVVPVLLVYCVFQRIAGSVSQGTER"
misc_feature
3554..3640
/gene="SCO6007"
/note="PS00402 Binding-protein-dependent transport systems
inner membrane cpm signature"
3935..3984
/note="hairpin loop with 23bp stem"
4128..4131
/note="possible RBS upstream of SC7B7.05"
4142..5353
/gene="SCO6008"
/note="synonym: SC7B7.05"
4142..5353
/gene="SCO6008"
/note="SC7B7.05, probable transcriptional repressor
protein, len 403 aa; similar to many eg. XYLIR ANATH Q4406
xylose repressor from Anaerocellum thermophilum (399 aa),
fasta scores; opt: 548 z-score: 767.3 E(): 0, 33.1%
identity in 393 aa overlap. Contains PS01125 ROK family
signature and helix-turn-helix motif at aa positions 29-50
(Score 2016, +6.05 SD)"
/codon_start=1
/transl_table=11
/product="probable transcriptional repressor protein"
/protein_id="CAA15785.1"
/db_xref="GI:2661680"
/db_xref="COA:O50502"
/db_xref="SPTREMBL:O50502"
/transl_table="METPQSSSLHRANLRRVVRVRLAGSLQFAETARTGLSAATV
SNIRREKDGTVETPTAGRRARSVLSGDAGIVGVDFGHTLRLVAVGNLAHQV
LAESPFLDASDDQFRAEOLVSRLEATGVDRAKIVAGLVGPGVDVDSGTLG
STALLPGGIRPAEELRGLRGPVVDNDANIGALGEINWGGGRRGVRDIAYIKVASG
VAGLVINGIKRPGGTAGEIGHIILDEAGPVCRGNCRLTFPAAARVVLPLQPS
HGVDLMEGVVRLRPGDPCRRVIADVGRHIGSVANLNLNPKRVVJGGDLAERAG
ELVGLPIRESVGYAIPSAARQLSVLPGALGSAEVLGALALASEMGDSLTLDGSAT
GALPAATPAFT"
misc_feature
4784..4867
/gene="SCO6008"
/note="PS01125 ROK family signature"
5462..5466
/note="possible RBS upstream of SC7B7.06"
5475..6590
/gene="SCO6009"
/note="synonym: SC7B7.06"
5475..6590
/gene="SCO6009"
/note="SC7B7.06, probable solute-binding protein of
transmembrane transport system, len: 403 aa; similar to
many eg. XYLIF ECOLI P37387 d-xylose-binding periplasmic
protein precursor (330 aa), fasta scores; opt: 619
z-score: 573.4 E(): 8.9e-25, 35.7% identity in 325 aa
overlap. Contains N-terminal signal sequence and
appropriately positioned PS00013 Prokaryotic membrane
lipoprotein lipid attachment site"
/codon_start=1
/transl_table=11
/product="solute-binding protein"
/protein_id="CAA15786.1"
/db_xref="GI:2661681"
/db_xref="SPTREMBL:O50503"
/transl_table="MNTMRRAAVAVATTAMAVSLAAGSKESDGSDDSSDAAKKG
DDIKVGLLPENQATRYEKDKPIEKVKELTNNGEVVYNAKQDASLQNVQDTM
Query Match 86.4%; Score 19; DB 1; Length 295150;
Best Local Similarity 100.0%; Pred. No. 5.9e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GGTGGAGTTCGAGGAGCTG 21
|||||
```

```
Db 233112 GGTGGAGTTCGAGGAGCTG 233130
RESULT 12
LOCUS
DEFINITION
AF110372 274 bp DNA linear INV 02-MAY-1999
Leishmania major clone LmV5-G1 putative glycoprotein gene, partial
cds.
ACCESSION
AF110372
VERSION
AF110372.1 GI:4731175
KEYWORDS
SOURCE
ORGANISM
Leishmania major
Leishmania major
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE
1 (bases 1 to 274)
AUTHORS
Ch'ang,L.Y. and Li,T.
TITLE
Trapping ORF-containing genomic fragments
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 274)
AUTHORS
Ch'ang,L.Y. and Li,T.
TITLE
Direct Submission
JOURNAL
Submitted (01-DEC-1998) Academia Sinica, Institute of Biomedical
Sciences, Rm 433, 128 Yen-Chiun-Yuan Road SEC 2, Taipei 11529,
R.O.C Taiwan
FEATURES
Location/Qualifiers
source
1..274
/organism="Leishmania major"
/mol_type="genomic DNA"
/db_xref="taxon:5664"
/clone="LmV5-G1"
<1..>274
/codon_start=2
/evidence="not experimental"
/product="putative glycoprotein"
/protein_id="AAD28377.1"
/db_xref="GI:4731176"
/transl_table="TEABRQKLKLEQHEGQFEARQRRVAEEKAQKRAEKABEA
EGELAATRRQKGELELQRAHEEBEKQRIEMVYRQREETQKKREI"
ORIGIN
Query Match 85.5%; Score 18.8; DB 3; Length 274;
Best Local Similarity 90.9%; Pred. No. 6.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
|||||
Db 167 AAGGGGAGTTCGAGGAGCTGC 188
RESULT 13
LOCUS
LEIGP9692A 704 bp DNA linear INV 27-JUL-1994
Leishmania major glycoprotein 96-92 (GP 96-92) gene, partial cds.
ACCESSION
M63109
VERSION
M63109.1 GI:159332
KEYWORDS
glycoprotein 96-92.
SOURCE
Leishmania major
ORGANISM
Leishmania major
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE
1 (bases 1 to 704)
AUTHORS
Handman,E., Barnett,L.D., Osborn,A.H., Goding,J.W. and Murray,P.J.
TITLE
Identification, characterisation and genomic cloning of a O-linked
N-acetylglucosamine-containing cytoplasmic Leishmania glycoprotein
JOURNAL
Mol. Biochem. Parasitol. 62 (1), 61-72 (1993)
MEDLINE
94158977
PUBMED
8114827
COMMENT
Original source text: Leishmania major DNA.
FEATURES
Location/Qualifiers
source
1..704
/organism="Leishmania major"
/mol_type="unassigned DNA"
```

```

/isoate="WR309"
/db_xref="taxon:5664"
/clone_lib="lambda Grl1"
1. .704
/gene="GP 96-92"
<1. .>704
/gene="GP 96-92"
/codon_start=3
/product="glycoprotein 96-92"
/protein_id="AAA20179.1"
/db_xref="GI:159333"
/translation="COAWLQVDSWSPVSAGEEBEKQRIEMVRKQREAAKKEEIQ
KQREIKRRAEIAERQKLEQHEHREQEARSRVAEKEKAQKKAABEAE
DELAATRRQELGELQHEHREQEARQRIEMVRKQREAAKKEEIQKQREIEIKRRK
AEIEARQKLEQHEHREQEARQRIEMVRKQREAAKKEAABEDELATRRQRK
GELBELQAREEABF"

ORIGIN
Query Match      85.5%; Score 18.8; DB 3; Length 704;
Best Local Similarity 90.9%; Pred. No. 6.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
||||| ||||| ||||| ||||| |||||

DB 339 AAGGGGAGTTGGAGGAGCTGC 360
```

```

RESULT 14
AY051720
LOCUS      2315 bp mRNA linear INV 27-AUG-2001
DEFINITION Drosophila melanogaster LD24980 full length cDNA.
ACCESSION AY051720
VERSION    AY051720.1 GI:15291750
KEYWORDS   FLI CDNA.
SOURCE     Drosophila melanogaster (fruit fly)
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 2315)
AUTHORS    Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
            Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R.,
            Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,
            Nunco,J., Pacleb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K.,
            Yu,C., Lewis,S.E., Rubin,G.M. and Celniker,S.
            Direct Submission
            Submitted (10-AUG-2001) Berkeley Drosophila Genome Project,
            Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, USA
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory
            Berkeley, CA 94720
            This clone was sequenced as part of a high-throughput process to
            sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
            Science 2000) . The sequence has been subjected to integrity checks
            for sequence accuracy, presence of a polyA tail and contiguity
            within 100 kb in the genome. Thus we believe the sequence to
            reflect accurately this particular cDNA clone. However, there are
            artifacts associated with the generation of cDNA clones that may
            have not been detected in our initial analyses such as internal
            priming, priming from contaminating genomic DNA, retained introns
            due to reverse transcription of unspliced precursor RNAs, and
            reverse transcriptase errors that result in single base changes.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our Web site
            (http://fruitfly.berkeley.edu) or send email to
            cdna@fruitfly.berkeley.edu.
FEATURES
            location/Qualifiers
            1. .2315
            /organism="Drosophila melanogaster"
            /mol_type="mRNA"
            /strain="Y; cn bw sp"
```

```

/db_xref="taxon:7227"
/map="16D7-16D7"
1. .2315
/gene="CG7846"
/note="alignment with genomic scaffold AB003507"
/db_xref="FLYBASE:FBgn0030877"
82. .1905
/gene="CG7846"
/note="Longest ORF"
/product="LD24980p"
/protein_id="AAK93144.1"
/db_xref="GI:15291751"
/translation="MQRSRASSTSSGRHPAPPICLPQAPQPLEAPKLIIVHPSQH
LRIGRAADNLPTLLHAYARRRGASDRPHDPLLPEDNVNSNSGLMVEFEORLV
VSRILQHCVDNKNRLRVATPPQQLAHFNRSQAQKVPAPSGQMADEPWLQREAPVLE
DDRIILRGAVDARNYDIHFPIQRGELNVHNEKGSLSQASMOHLERINWVAIEERLKIP
LRKLGTCAVLVNDVYVRHLRFVTLRLRGLFRRCFLVQDSVASTFGAGYGCYV
VDIGAOKTSIACIDIGISQIDARVRLSYGGDLQVLLLLLRKCGFPYREGNVQESYV
DAHLDELKEKFCNLNASVCGAQEKHNLKHNQOMLYTIOVGDEALMAPLALFHE
LLNITGRKAVFTQAAVDQDCDFDAEYKLTGRKNGVRGGDILQLSTASAGQPR
POLPVTADDELI VVDQDETISNCQSIGAQTAGQMNSGCHNGQGLVLPDQAIL
QSNRLSYETKRMFGSILVSGSAKLPLAALWLEQRIISQOVQSEVNVLIKGMADGM
VAVKGAATMSVLESARELWISQNDWQRHGLRVLRERSPLW"

ORIGIN
Query Match      85.5%; Score 18.8; DB 3; Length 2315;
Best Local Similarity 90.9%; Pred. No. 6.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
||||| ||||| ||||| ||||| |||||

DB 358 ATGTTGGAGTTGGAGGAGCAGC 379

RESULT 15
AY059441
LOCUS      2905 bp mRNA linear INV 12-OCT-2001
DEFINITION Drosophila melanogaster GH08948 full length cDNA.
ACCESSION AY059441
VERSION    AY059441.1 GI:16076857
KEYWORDS   FLI CDNA.
SOURCE     Drosophila melanogaster (fruit fly)
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 2905)
AUTHORS    Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
            Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R.,
            Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,
            Nunco,J., Pacleb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K.,
            Yu,C., Lewis,S.E., Rubin,G.M. and Celniker,S.
            Direct Submission
            Submitted (08-OCT-2001) Berkeley Drosophila Genome Project,
            Lawrence Berkeley National Laboratory, One Cyclotron Road,
            Berkeley, CA 94720, USA
COMMENT     Sequence submitted by:
            Berkeley Drosophila Genome Project
            Lawrence Berkeley National Laboratory
            Berkeley, CA 94720
            This clone was sequenced as part of a high-throughput process to
            sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
            Science 2000) . The sequence has been subjected to integrity checks
            for sequence accuracy, presence of a polyA tail and contiguity
            within 100 kb in the genome. Thus we believe the sequence to
            reflect accurately this particular cDNA clone. However, there are
            artifacts associated with the generation of cDNA clones that may
            have not been detected in our initial analyses such as internal
            priming, priming from contaminating genomic DNA, retained introns
            due to reverse transcription of unspliced precursor RNAs, and
            reverse transcriptase errors that result in single base changes.
```

For further information about this sequence, including its location and relationship to other sequences, please visit our Web site (<http://fruitfly.berkeley.edu>) or send email to cdna@fruitfly.berkeley.edu.

FEATURES

```

source
1..2905
    Location/Qualifiers
        organism="Drosophila melanogaster"
        mol_type="mRNA"
        strain="y; cn bw sp"
        db_xref="taxon:7227"
        map="61E1-61E2"
1..2905
    /gene="Glut1"
    /note="alignment with genomic scaffold AE003470. gene does
not completely overlap longest ORF"
    /db_xref="FLYBASE:FBgn0025593"
953..2755
    /gene="Glut1"
    /note="Longest ORF"
    /codon_start=1
    /product="GH08948p"
    /protein_id="AAU1347.1"
    /db_xref="GI:16076858"
    /db_xref="FLYBASE:FBgn0025593"
    /translation="MSNLSDSDYDVFGRRKTPPTTSTQLKLTSPVPELRPEAYR
GANINNGRFVYRH3FSYAPKEQRHSSRRDRRESRLRCHGEDEATLQLLLDLQKQSV
MSNLAKLDELQGRDRLHETITVALCEITQLQELTKSVESQSEVSEVRKDMVAIKH
ELDTVQQVKEEIBELREYVDRLEBETHRRKLRLLEQGLTFELTYSIFSAYLGMQLQGY
NTGVINAPEKNIENFKVDYKDRYGEDISEEFIQQLYSVAVSIFAIGMLGFGSGWM
ANRFGKGGILLNNVLGIAGACLMGPTKVSHSYEMILGILGTNEGWPILLGLAICPAILOLI
SETAPLNLRGGLGVNOLAVTVGLLLSQVLGIEQILGTNEGWPILLGLAICPAILOLI
LLPVCPSPYLLITKQWEEEARLARLRRLRASCVSVEDIEEMRAERAQOQSESHISTM
ELICSPFLRPFLIGIVMQLSQQFSGINAVFYISTLSFSSGLTERSAKFATIGIGAI
MVYMTLVSIPMDRTGRTHLHLYGLGGMFIFSIPTISFLIKASIVVSRILESCSCSC
RVMPANVNAKMPASLGLHLEVPVPFPFSDLHMTLKS"

```

ORIGIN

```

Query Match      85.5%; Score 18.8; DB 3; Length 2905;
Best Local Similarity 90.9%; Pred. No. 6.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  AAGGTGGAGTTCGAGGAGCTGC 22
      ||||| ||||| ||||| |||||
Db     1457 AAGGAGGAGATCGAGGAGCTGC 1478

```

Search completed: July 18, 2004, 13:33:22
Job time : 211.026 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: July 18, 2004, 11:40:25 ; Search time 208.026 Seconds
(without alignments)
4583.781 Million cell updates/sec

Title: US-10-051-769-6
Perfect score: 22
Sequence: 1 gtggaagccgcgttgtaactcc 22

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	22	100.0	243412	2	AC127769	AC127769 Rattus no
C 2	20.4	92.7	2623	9	BC012186	BC012186 Homo sapi
C 3	20.4	92.7	2639	6	AX747159	AX747159 Sequence
C 4	20.4	92.7	2639	9	AK091746	AK091746 Homo sapi
C 5	20.4	92.7	3847	6	BD169701	BD169701 Human gli
C 6	20.4	92.7	3856	9	BC028375	BC028375 Homo sapi
7	20.4	92.7	49052	9	AL359836	AL359836 Human DNA
8	20.4	92.7	68196	2	AC108407	AC108407 Mus muscu
C 9	20.4	92.7	120578	9	AC005887	AC005887 citb 173
C 10	20.4	92.7	221524	10	AC139040	AC139040 Mus_muscu
C 11	18.8	85.5	960	6	AR070450	AR070450 Sequence
C 12	17.8	80.9	110000	2	LMFCHR36_16	Continuation (17 o
C 13	17.8	80.9	192579	2	AC141584	AC141584 Rattus no
C 14	17.4	79.1	1935	6	BD182190	BD182190 Polynucle
C 15	17.4	79.1	1935	6	BD188593	BD188593 Polynucle
C 16	17.4	79.1	8056	6	BD182188	BD182188 Polynucle
C 17	17.4	79.1	8056	6	BD188591	BD188591 Polynucle
C 18	17.4	79.1	9901	6	BD182263	BD182263 Polynucle
C 19	17.4	79.1	9901	6	BD188666	BD188666 Polynucle
C 20	17.2	78.2	66	6	AX612540	AX612540 Sequence
C 21	17.2	78.2	210	1	AY187875	AY187875 Escherich
C 22	17.2	78.2	358	5	AF266215	AF266215 Gillichth
C 23	17.2	78.2	618	6	E12430	E12430 DNA encodin
C 24	17.2	78.2	681	6	AR065638	AR065638 Sequence
C 25	17.2	78.2	681	6	AR096680	AR096680 Sequence
C 26	17.2	78.2	711	4	SSU81236	U81236 Sus scrofa
C 27	17.2	78.2	723	9	HSU81237	U81237 Human von W
C 28	17.2	78.2	750	6	A76864	A76864 Sequence 2
C 29	17.2	78.2	1076	4	AF061063	AF061063 Tupaia gl
C 30	17.2	78.2	1076	4	TGU31624	U31624 Tupaia glis
C 31	17.2	78.2	1155	4	DKA427361	AJ427361 Dasyptus k
C 32	17.2	78.2	1203	4	AF076480	AF076480 Chaetophr
C 33	17.2	78.2	1220	9	TBA410296	AJ410296 Tarsius b
C 34	17.2	78.2	1230	4	MSU97535	U97535 Manis sp. v
C 35	17.2	78.2	1231	4	ZPI427365	AJ427365 Zaedyus p
C 36	17.2	78.2	1232	4	AF226849	AF226849 Megaptera
C 37	17.2	78.2	1233	4	AF304109	AF304109 Megaptera
C 38	17.2	78.2	1233	4	DNO278158	AJ278158 Dasyptus n
C 39	17.2	78.2	1233	10	PTY251144	AJ251144 Petromus
C 40	17.2	78.2	1234	4	AF108834	AF108834 Phytseter
C 41	17.2	78.2	1236	4	TMA427362	AJ427362 Tolypsteu
C 42	17.2	78.2	1238	9	HLA410300	AJ410300 Hylobates
C 43	17.2	78.2	1239	4	ESE427364	AJ427364 Euphractu
C 44	17.2	78.2	1242	4	OPAJ4672	AJ224672 Ochotona
C 45	17.2	78.2	1242	4	S78431	S78431 von Willebr

ALIGNMENTS

RESULT 1
AC127769/c
LOCUS AC127769 243412 bp DNA linear HTG 19-NOV-2002
DEFINITION Rattus norvegicus clone CH230-157C16, WORKING DRAFT SEQUENCE, 3
unordered pieces.
ACCESSION AC127769.3 GI:25079525
VERSION HTG; HTGS PHASR1; HTGS DRAFT; HTGS_FULLTOP.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 243412)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,

Allen, C., Allen, H., Alshrocks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Curch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Guebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, M., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Loreshehwa, L., Loulseghe, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mahoney, S., McLeod, M.P., McNeill, T.Z., Meenen, B., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackelmech, O., Okwuon, G., Olarpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quatroz, J., Rachlin, E., Reeves, R., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajis, D., Snead, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umami, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

COMMENT

Unpublished

2 (bases 1 to 243412)

Worley, K.C.

Direct Submission

Submitted (19-JUL-2002)

Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 243412)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (19-NOV-2002)

Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Nov 19, 2002 this sequence version replaced gi:23269487.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome

shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: G2SV

Center clone name: CH230-157C16

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 220496 bases at least Q40

Consensus quality: 222898 bases at least Q30

Consensus quality: 224623 bases at least Q20

Estimated insert size: 224491; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

consists of 3 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

* 1 11919: contig of 11919 bp in length

* 11920 12019: gap of unknown length

* 12020 18301: contig of 6282 bp in length

* 18302 18401: gap of unknown length

* 18402 243412: contig of 225011 bp in length.

FEATURES

source

1. 243412

/organism="Rattus norvegicus"

/mol_type="Genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-157C16"

1. 1535

/note="wgs_contig"

12020. 13436

/note="wgs_contig"

17209. 18301

/note="wgs_contig"

18402. 19442

/note="wgs_contig"

25075. 27269

/note="wgs_contig"

238399. 242288

/note="wgs_contig"

242339. 243412

/note="wgs_contig"

ORIGIN

Query Match 100.0%; Score 22; DB 2; Length 243412;

Best Local Similarity 100.0%; Pred. No. 24;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTGGAAGCGCGGTGTACTCC 22

DB 91073 GTGGAAGCGCGGTGTACTCC 91052

RESULT 2

BC012186/c

LOCUS BC012186

DEFINITION Homo sapiens, clone IMAGE:4564853, mRNA.

ACCESSION BC012186

VERSION BC012186.1 GI:15082555

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Homo sapiens

PRI 06-AUG-2001

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2623)
Strausberg,R.
Direct Submission
Submitted (02-AUG-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapsb@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabh, Parvaneh Saeedi, Jacqueline Schein, Duane Smalus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 29 Row: h Column: 17
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Genomescan gene prediction.
analysis, Location/Qualifiers
1. .2623
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4564853"
/tissue_type="Kidney, renal cell adenocarcinoma"
/clone_lib="NIH MGC 14"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4564853"
/tissue_type="Kidney, renal cell adenocarcinoma"
/clone_lib="NIH MGC 14"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
AX747159/c
LOCUS AX747159 2639 bp mRNA linear PAT 20-JUN-2003
DEFINITION Sequence 684 from Patent EP1308459.
ACCESSION AX747159
VERSION AX747159.1 GI:32131547
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R., Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and Masuho,Y.
Full-length cDNA sequences
Patent: EP 1308459-A 684 07-MAY-2003;
Helix Research Institute (JP) ; Research Association for

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 Isogai,T., Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J.I., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2639)
Isogai,T. and Yamamoto,J.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel.81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB.
RAB; annotation: HRI and RAB.
Location/Qualifiers
1. .2639
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HHDP2008816"
/cell_type="dermal papilla cells (HDPC)"
/clone_lib="HHDP2"
/note="Cloning vector: pME18SFL3-primary culture, dermal papilla cells"
GTGGAAGCCCGCTGTGACTCC 22
|||||||
822 GTGGAAGCCCGCTGTGACTCC 801

Query Match 92.7%; Score 20.4; DB 9; Length 2639;
Best Local Similarity 95.5%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GTGGAAGCCCGCTGTGACTCC 22
|||||||
822 GTGGAAGCCCGCTGTGACTCC 801

Query Match 92.7%; Score 20.4; DB 9; Length 2639;
Best Local Similarity 95.5%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GTGGAAGCCCGCTGTGACTCC 22
|||||||
822 GTGGAAGCCCGCTGTGACTCC 801

```

RESULT 5
LOCUS      BD169701/c
DEFINITION Human glioma antigen and method of preparing the same.
ACCESSION  BD169701
VERSION     BD169701.1 GI:27875513
KEYWORDS   WO 02055695-A/6.
SOURCE      Homo sapiens (human)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3847)
AUTHORS     Toda,M., Kawakami,Y., Kawase,T. and Iizuka,Y.
TITLE       Human glioma antigen and method of preparing the same
JOURNAL     Patent: WO 02055695-A 6 18-JUL-2002;
            KEIO UNIVERSITY,WASHIRO TODA,FUTAKA KAWAKAMI,TAKESHI KAWASE,
            YUKIHIKO IIZUKA
COMMENT     OS Homo sapiens (human)
            PN WO 02055695-A/6
            PD 18-JUL-2002
            PE 30-NOV-2001 WO 2001JP010505
            PR 09-JAN-2001 JP 01P 001965
            PT MASAHIRO TODA,FUTAKA KAWAKAMI,TAKESHI KAWASE,YUKIHIKO IIZUKA
            CI C12N15/12,C12N5/10,A01K67/027,A61K31/711,A61K38/00,A61K39/00,
            PC A61K39/395,
            PC A61K45/00,A61K48/00,A61P35/00,C07K14/47,C07K16/18,C07K19/00,
            PC C12Q1/02,
            PC C12Q1/68
            CC Human glioma antigen and method of preparing the same FH Key
FEATURES   source
            FT CDS
               Location/Qualifiers
               1..3847
               /organism="Homo sapiens"
               /mol_type="genomic DNA"
               /db_xref="taxon:9606"
ORIGIN
Query Match 92.7%; Score 20.4; DB 6; Length 3847;
Best Local Similarity 95.5%; Pred. No. 2.7e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GTGAAGCCCGTGTACTCC 22
Db 30 GTGAAGCCCGTGTACTCC 9

RESULT 6
LOCUS      BC028375/c
DEFINITION Homo sapiens hypothetical protein LOC118987, mRNA (cDNA clone
ACCESSION  BC028375
VERSION     BC028375.1 GI:22382223
KEYWORDS   MGC.
SOURCE      Homo sapiens (human)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3856)
AUTHORS     Strausberg,R., M., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Narusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Scapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.H.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Frange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smallos,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
            Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
            Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
            22388257
            MEDLINE
            PUBMED
            12477932
            REFERENCE
            2 (bases 1 to 3856)
            Strausberg,R.
            Direct Submission
            Submitted (23-APR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
            NIH-MGC Project URL: http://mgc.nci.nih.gov
            Contact: MGC help desk
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbio.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAC Plate: 34 Row: m Column: 21
            This clone was selected for full length sequencing because it
            passed the following selection criteria: GenomeScan gene
            prediction.
            Location/Qualifiers
            1..3856
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="MGC:27107 IMAGE:4837939"
            /tissue type="Testis"
            /clone_lib="NIH MGC_97"
            /lab_host="DH10B"
            /note="Vector: pBluescript"
            1..3856
            /gene="LOC118987"
            /db_xref="LocusID:118987"
            203..3667
            /codon_start=1
            /product="hypothetical protein LOC118987"
            /protein_id="AAH28375.1"
            /db_xref="GI:22382224"
            /db_xref="LocusID:118987"
            /translation="MGILLMLASAVLGSFLLTAQFFLLYRQPEPPADRAARAGE
            FRIKVPGLLLREYLYGGRDEPFGAEPGAFTAPPTAPPTTRETCTYFNATI
            LFLREURDALTALTRWTKIKVEFELLQTKTAGRLLEGLSLDRDVLGTVFPFKTI
            RLPRVPSATGEDPGEPAALPAACEELAFEAVEYNGGFLHAIIDLVLVFGKSAYL
            FKLRSVVGRLRLVFPVTFWFFSFEVDPLIDFEVRSQPEGSPMPLQTLIIYNQK
            KIIRKHTLPNFKIRFPFPYQTLQGFDEDEHHIQOVALTEGRLKVLTECSRLL
            IFGSYDEANVHCTLELSSSWEEKQSSIKTVELIKGNLQSLGTLRLVOSTGVAG
            HVIIETVAPNSPAALDQGRDRLAIGGVKTIQVLTQVLKLIKAGDRVLVYERPVG
            QSNOGVLQDNFGQLEENFLSSSCQSGTEEAAGLTVDTSRELDSEDFEDLASDVRQ
            NEFKDEAQSLSHSKPRVPTLISIKPLGAIISVLNKLAVGSHPLPPKIQSKDGNKPP
            LKTSIEDTPAQVSKPTQSGAKPPVPPRPAQKVLPSADAPNPAEPDLVLEKPKVPP
            PVLVDKSAEKQAKNVDAIDDAAPFLAKQAEVAKDVTSETSCPTKSSDDROTWESS
            EILVNLKIGKWTTRASCLDIEACHRYLALWCRDPFKLGGICLGHVSLKLEDA
            LGCLATNTEVLSKRLKLEAPSKAIVTRALNLSMQKGNFKCYGDIITHFKYKE
            GESDHHVTVNEKEKHEHLEEVSVLPKEQFVGQMLTENKHSFQDTQNPWTWCDY
            CKKRWTKAASQCMFCAYVCHKKQCEKLAETSVCGATDRIDRTLKNLRLEGQETLL
            GLPRVDAEASKSVNKTGLTRHINTSSRLNLRQVSKTRLSPEGFDLVEPSPKHTP
            NTSNDESGTEVCCPNSPSKRGNSTGIKLVRKEGGLDSDSVFAVKEIGRDLYRGLPTE

```

ERIOKLEFMDLKLONEIDOLEHNSLVREEKETTDRKKSLLSALAKSGERLOALT
 LLMHYRAGIEDIETLESLSLDQSKSKISKYTDTEEDLDNEISQIDSPSSISDD
 LFGPSESV"

misc_feature

1292..11546
 /note="PDZ; Region: Domain present in PSD-95, Dlg, and ZO-1/2. Also called DHR (Dlg homologous region) or GLGF (relatively well conserved tetrapeptide in these domains). Some PDZs have been shown to bind C-terminal polypeptides"
 /db_xref="CDD:smar00228"
 2723..2851
 /note="DAG PE-bind; Region: Phorbol esters/diacylglycerol binding domain (C1 domain). This domain is also known as the Protein kinase C conserved region 1 (C1) domain"
 /db_xref="CDD:pfam00130"

misc_feature

2723..2851
 /note="DAG PE-bind; Region: Phorbol esters/diacylglycerol binding domain (C1 domain). This domain is also known as the Protein kinase C conserved region 1 (C1) domain"
 /db_xref="CDD:pfam00130"

ORIGIN

Query Match 92.7%; Score 20.4; DB 9; Length 3856;
 Best Local Similarity 95.5%; Pred. No. 2.7e+02;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGGAAGCCCGCTTGTACTCC 22

|||||

Db 811 GTGGAAGCCCGCTTGTACTCC 790

RESULT 7

AL359836 49052 bp DNA linear PRI 21-DEC-2001
 LOCUS Human DNA sequence from clone RP11-389E6 on chromosome 10, complete
 DEFINITION sequence.

ACCESSION

AL359836

VERSION

AL359836.16 GI:17977720

KEYWORDS

HTG.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. Smith, M.

REFERENCE

1 (bases 1 to 49052)

AUTHORS

Direct Submission

Submitted (21-DEC-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

On Dec 23, 2001 this sequence version replaced gi:17384082.

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30). An attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr10

RP11-389E6 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-389E6 It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone CTA-109P11 is at 47053 in this sequence.
 The true right end of clone RP11-129M16 is at 2000 in this sequence.

FEATURES

source
 1..49052
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone="RP11-389E6"
 /clone_lib="RPCI-11.2"

ORIGIN

Query Match 92.7%; Score 20.4; DB 9; Length 49052;
 Best Local Similarity 95.5%; Pred. No. 1.6e+02;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GTGGAAGCCCGCTTGTACTCC 22
 |||||

Db 29524 GTGGAAGCCCGCTTGTACTCC 29545

RESULT 8

AC108407 68196 bp DNA linear HTG 27-JAN-2002
 LOCUS Mus musculus clone RP24-422P10, LOW-PASS SEQUENCE SAMPLING.
 DEFINITION AC108407

AC108407.1 GI:18377216

ACCESSION

HTG; HTGS PHASE0.

KEYWORDS

Mus musculus (house mouse)

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. Birren, B., Linton, L., Nusbaum, C. and Lander, E.

REFERENCE

1 (bases 1 to 68196)

AUTHORS

Unpublished

TITLE

Mus musculus, clone RP24-422P10

JOURNAL

2 (bases 1 to 68196)

REFERENCE

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-pierre, N., Hagos, B., Horton, L., Hulme, W., Illiev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

Direct Submission

JOURNAL

Submitted (27-JAN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L18881

Center clone name: 422_P_10

* NOTE: This record contains 85 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 714: contig of 714 bp in length
* 715 814: gap of 100 bp
* 815 1520: contig of 706 bp in length
* 1521 1620: gap of 100 bp
* 1621 2323: contig of 703 bp in length
* 2324 2423: gap of 100 bp
* 2424 3134: contig of 711 bp in length
* 3135 3234: gap of 100 bp
* 3235 3937: contig of 703 bp in length
* 3938 4037: gap of 100 bp
* 4038 4739: contig of 702 bp in length
* 4740 4839: gap of 100 bp
* 4840 5545: contig of 706 bp in length
* 5546 5645: gap of 100 bp
* 5646 6352: contig of 707 bp in length
* 6353 6452: gap of 100 bp
* 6453 7172: contig of 720 bp in length
* 7173 7272: gap of 100 bp
* 7273 7996: contig of 724 bp in length
* 7997 8096: gap of 100 bp
* 8097 8814: contig of 718 bp in length
* 8815 8914: gap of 100 bp
* 8915 9629: contig of 715 bp in length
* 9630 9729: gap of 100 bp
* 9730 10417: contig of 688 bp in length
* 10418 10517: gap of 100 bp
* 10518 11223: contig of 706 bp in length
* 11224 11323: gap of 100 bp
* 11324 12014: contig of 691 bp in length
* 12015 12114: gap of 100 bp
* 12115 12816: contig of 702 bp in length
* 12817 12916: gap of 100 bp
* 12917 13616: contig of 700 bp in length
* 13617 13716: gap of 100 bp
* 13717 14422: contig of 706 bp in length
* 14423 14522: gap of 100 bp
* 14523 15234: contig of 712 bp in length
* 15235 15334: gap of 100 bp
* 15335 16043: contig of 709 bp in length
* 16044 16143: gap of 100 bp
* 16144 16859: contig of 716 bp in length
* 16860 16959: gap of 100 bp
* 16960 17669: contig of 710 bp in length
* 17670 17769: gap of 100 bp
* 17770 18460: contig of 691 bp in length
* 18461 18560: gap of 100 bp
* 18561 19269: contig of 709 bp in length
* 19270 19369: gap of 100 bp
* 19370 20070: contig of 701 bp in length
* 20071 20170: gap of 100 bp
* 20171 20874: contig of 704 bp in length
* 20875 20974: gap of 100 bp
* 20975 21684: contig of 710 bp in length
* 21685 21784: gap of 100 bp
* 21785 22491: contig of 707 bp in length
* 22492 22591: gap of 100 bp
* 22592 23299: contig of 708 bp in length
* 23300 23399: gap of 100 bp
* 23400 24086: contig of 687 bp in length
* 24087 24186: gap of 100 bp

24187 24901: contig of 715 bp in length
* 24902 25001: gap of 100 bp
* 25002 25714: contig of 713 bp in length
* 25715 25814: gap of 100 bp
* 25815 26528: contig of 714 bp in length
* 26529 26628: gap of 100 bp
* 26629 27358: contig of 730 bp in length
* 27359 27458: gap of 100 bp
* 27459 28159: contig of 701 bp in length
* 28160 28259: gap of 100 bp
* 28260 28966: contig of 707 bp in length
* 28967 29066: gap of 100 bp
* 29067 29765: contig of 699 bp in length
* 29766 29865: gap of 100 bp
* 29866 30577: contig of 712 bp in length
* 30578 30677: gap of 100 bp
* 30678 31381: contig of 704 bp in length
* 31382 31481: gap of 100 bp
* 31482 32192: contig of 711 bp in length
* 32193 32292: gap of 100 bp
* 32293 32974: contig of 682 bp in length
* 32975 33074: gap of 100 bp
* 33075 33760: contig of 686 bp in length
* 33761 33860: gap of 100 bp
* 33861 34574: contig of 714 bp in length
* 34575 34674: gap of 100 bp
* 34675 35378: contig of 704 bp in length
* 35379 35478: gap of 100 bp
* 35479 36194: contig of 716 bp in length
* 36195 36294: gap of 100 bp
* 36295 36973: contig of 679 bp in length
* 36974 37073: gap of 100 bp
* 37074 37774: contig of 701 bp in length
* 37775 37874: gap of 100 bp
* 37875 38586: contig of 712 bp in length
* 38587 38686: gap of 100 bp
* 38687 39388: contig of 702 bp in length
* 39389 39488: gap of 100 bp
* 39489 40195: contig of 707 bp in length
* 40196 40295: gap of 100 bp
* 40296 41013: contig of 718 bp in length
* 41014 41113: gap of 100 bp
* 41114 41828: contig of 715 bp in length
* 41829 41928: gap of 100 bp
* 41929 42643: contig of 715 bp in length
* 42644 42743: gap of 100 bp
* 42744 43458: contig of 715 bp in length
* 43459 43558: gap of 100 bp
* 43559 44242: contig of 684 bp in length
* 44243 44342: gap of 100 bp
* 44343 45043: contig of 701 bp in length
* 45044 45143: gap of 100 bp
* 45144 45856: contig of 713 bp in length
* 45857 45956: gap of 100 bp
* 45957 46654: contig of 698 bp in length
* 46655 46754: gap of 100 bp
* 46755 47449: contig of 695 bp in length
* 47450 47549: gap of 100 bp
* 47550 48258: contig of 709 bp in length
* 48259 48358: gap of 100 bp
* 48359 49070: contig of 712 bp in length
* 49071 49170: gap of 100 bp
* 49171 49886: contig of 716 bp in length
* 49887 49986: gap of 100 bp
* 49987 50673: contig of 687 bp in length
* 50674 50773: gap of 100 bp
* 50774 51472: contig of 699 bp in length
* 51473 51572: gap of 100 bp
* 51573 52288: contig of 716 bp in length
* 52289 52388: gap of 100 bp
* 52389 53075: contig of 687 bp in length
* 53076 53175: gap of 100 bp
* 53176 53879: contig of 704 bp in length

```

* 53880 53979: gap of 100 bp
* 53980 54675: contig of 696 bp in length
* 54775: gap of 100 bp
* 54776 55477: contig of 702 bp in length
* 55478 55577: gap of 100 bp

Query Match
Best Local Similarity 92.7%; Score 20.4; DB 2; Length 68196;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGGAAGCCCGCTGTACTCC 22
||||| |||||||
Db 17262 GTGGAACCCCGTGTACTCC 17283

RESULT 9
AC005887/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT
FEATURES
source
1. 120578
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/map="10q25"
/clone_lib="CIT987SK-1173112"

ORIGIN
Query Match
Best Local Similarity 92.7%; Score 20.4; DB 9; Length 120578;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGGAAGCCCGCTGTACTCC 22
||||| |||||||
Db 54219 GTGGAAGCCCGCTGTACTCC 54198

RESULT 10
AC139040/c

```

```

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Mus musculus (house mouse)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 221524)
Mus musculus chromosome 19, clone RP23-292H20
Unpublished
2 (bases 1 to 221524)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Menes,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-JAN-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 221524)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Menes,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (31-MAY-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 221524)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,

```

Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vasiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (16-JUL-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 16, 2003 this sequence version replaced gi:31249817.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L29063
Center clone name: 292_H_20

FEATURES

source

Location/Qualifiers

1..221524
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="19"
/map="19"
/clone="RP23-292H20"
/clone_lib="RPCI-23 Female Mouse BAC"
1..7046
/note="wgs end extension"
clone end:SP6"
1234..1239
/note="<30 qual SNGL region"
complement(1312..1743)
/rpt_family="MLTIH1"
complement(1981..2123)
/rpt_family="MER117"
2143..2511
/rpt_family="ORR1D"
2897..2955
/rpt_family="(CA)n"
3063..3114
/rpt_family="(TG)n"
3141..3217
/rpt_family="(CA)n"
complement(3235..3402)
/rpt_family="B4A"
complement(3447..3584)
/rpt_family="B1_MM"
complement(3456..3690)
/rpt_family="ID_B1"
3724..3752
/rpt_family="(A)n"
complement(4335..4351)
/rpt_family="B3A"
4352..4408
/rpt_family="(TG)n"
complement(4409..4625)
/rpt_family="B3A"
complement(4701..4799)
/rpt_family="MIR"
complement(5000..5112)
/rpt_family="B1F"
5268..5431
/rpt_family="ORR1D"
5782..5870
/rpt_family="B4"

repeat_region
/rpt_family="B1_MM"
6447..6481
/rpt_family="(TC)n"
6481..6516
/rpt_family="(CA)n"
7047..7052
/note="clone boundary"
clone end:SP6
site:EcoRI"
7196..7246
/rpt_family="(GAAA)n"
complement(7364..7656)
/rpt_family="ORR1C"
8336..8484
/rpt_family="B1_MM"
8968..9042
/rpt_family="L2"
complement(9139..9332)
/rpt_family="MIR"
complement(9553..9742)
/rpt_family="B2_MM2"
complement(9847..10055)
/rpt_family="B3"
11638..11844
/rpt_family="B3"
12218..12494
/rpt_family="RMER14"
complement(13731..13993)
/rpt_family="B4"
14922..14986
/rpt_family="MIR"
15105..15218
/rpt_family="L1MB3"
15345..15509
/rpt_family="L1MC4"
complement(15704..15731)
/rpt_family="ID_B1"
15732..15849
/rpt_family="ORR1D"
15850..15874
/rpt_family="(TGGGGG)n"
15875..16100
/rpt_family="ORR1D"
complement(16101..16251)
/rpt_family="ID_B1"
17479..17513
/rpt_family="(CA)n"
17626..17653
/rpt_family="(CCCA)n"
18382..18451
/rpt_family="ID4"
complement(18592..18730)
/rpt_family="RMER10A"

Query Match 92.7%; Score 20.4; DB 10; Length 221524;
Best Local Similarity 95.5%; Pred No. 1,1e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGGAAGCCGCGTTGTACTCC 22
|||||
Db 43076 GTGGAACCGCGTTGTACTCC 43055

RESULT 11
AR070450/c

LOCUS AR070450
DEFINITION Sequence 1 from patent US 5900476.
ACCESSION AR070450
VERSION AR070450.1 GI:7221338
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unknown.

linear
960 bp
PAT 18-FEB-2000

Unclassified.
REFERENCE 1 (bases 1 to 960)
AUTHORS Ruggieri,Z.M. and Ware,J.L.
TITLE Therapeutic domains of van Willebrand factor
JOURNAL Patent: US 5900476-A 1 04-MAY-1999;
FEATURES Location/Qualifiers
source 1..960
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 85.5%; Score 18.8; DB 6; Length 960;
Best Local Similarity 90.9%; Pred. No. 1.7e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTGGAAGCCCGCTGTGTACTCC 22
||||| ||||| ||||| ||||| |||||
Db 399 GTGGAAGCCCGCTGTGTACTCC 378

RESULT 12
LMFLCHR36_16/c
WPCOMMENT
Sequence split into 36 fragments LOCUS LMFLCHR36 Accession AL499624

Fragment Name	Begin	End
LMFLCHR36_00	1	110000
LMFLCHR36_01	100001	210000
LMFLCHR36_02	200001	310000
LMFLCHR36_03	300001	410000
LMFLCHR36_04	400001	510000
LMFLCHR36_05	500001	610000
LMFLCHR36_06	600001	710000
LMFLCHR36_07	700001	810000
LMFLCHR36_08	800001	910000
LMFLCHR36_09	900001	1010000
LMFLCHR36_10	1000001	1110000
LMFLCHR36_11	1100001	1210000
LMFLCHR36_12	1200001	1310000
LMFLCHR36_13	1300001	1410000
LMFLCHR36_14	1400001	1510000
LMFLCHR36_15	1500001	1610000
LMFLCHR36_16	1600001	1710000
LMFLCHR36_17	1700001	1810000
LMFLCHR36_18	1800001	1910000
LMFLCHR36_19	1900001	2010000
LMFLCHR36_20	2000001	2110000
LMFLCHR36_21	2100001	2210000
LMFLCHR36_22	2200001	2310000
LMFLCHR36_23	2300001	2410000
LMFLCHR36_24	2400001	2510000
LMFLCHR36_25	2500001	2610000
LMFLCHR36_26	2600001	2710000
LMFLCHR36_27	2700001	2810000
LMFLCHR36_28	2800001	2910000
LMFLCHR36_29	2900001	3010000
LMFLCHR36_30	3000001	3110000
LMFLCHR36_31	3100001	3210000
LMFLCHR36_32	3200001	3310000
LMFLCHR36_33	3300001	3410000
LMFLCHR36_34	3400001	3510000
LMFLCHR36_35	3500001	3529852

Continuation (17 of 36) of LMFLCHR36 from base 1600001 (AL499624 Leishmania major chromosome)

Query Match 80.9%; Score 17.8; DB 2; Length 110000;
Best Local Similarity 90.5%; Pred. No. 1.7e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCGAAGCCCGCTGTGTACTCC 22
||||| ||||| ||||| ||||| |||||
Db 21296 TCGAAGCCCGCTGTGTACTCC 21276

RESULT 13

AC141584/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-247A23, *** SEQUENCING IN PROGRESS
***, 72 unordered pieces.
AC141584.2 GI:29123824
VERSION HTG; HTGS_PHASE1.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Rattus.
REFERENCE 1 (bases 1 to 192579)
AUTHORS Muzny,D.,Marie., Metzker,M.,Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Georgatopis,E., Geer,K., Gill,R., Grady,M., Guerra,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,X., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpthy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A.,
Miner,G., Minja,E., Montanayor,J., Moore,S., Morgan,M., Morris,K.,
Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
Newton,N., Nguyen,N., Norris,S., Nwaokemele,O., Okwuonu,G.,
Olarunpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A.,
Popovic,D., Primus,E., Pu,L., Puazo,M., Quiroz,J., Rachlin,B.,
Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
Sisson,I., Sitter,C.D., Smaj,S., Sneed,A., Sodergren,E.,
Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villaseana,D.,
Waldfon,L., Walker,B., Wang,Q., Wang,S., Warren,J.,
Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
Holt,R.A., Smith,H.O., Weinstein,G. and Gibbs,R.A.
Direct Submission
Unpublished
REFERENCE 2 (bases 1 to 192579)
AUTHORS Worley,K.C.
TITLE
JOURNAL Submitted (18-MAR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 192579)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (19-MAR-2003) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Mar 19, 2003 this sequence version replaced gi:29029160.

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KEEH
Center clone name: CH230-247A23
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 164290 bases at least Q40
Consensus quality: 173116 bases at least Q30
Consensus quality: 179982 bases at least Q20
Estimated insert size: 176498; sum-of-contigs estimation
Quality coverage: 2x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 72 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1312: contig of 1312 bp in length
* 1313 1412: gap of unknown length
* 1413 2496: contig of 1084 bp in length
* 2497 2596: gap of unknown length
* 2597 3961: contig of 1365 bp in length
* 3962 4061: gap of unknown length
* 4062 5364: contig of 1303 bp in length
* 5365 5464: gap of unknown length
* 5465 6713: contig of 1249 bp in length
* 6714 7954: contig of 1141 bp in length
* 7955 8054: gap of unknown length
* 8055 9409: contig of 1354 bp in length
* 9409 9509: gap of unknown length
* 9509 10857: contig of 1348 bp in length
* 10857 12242: contig of 1286 bp in length
* 12243 12342: gap of unknown length
* 12343 13387: contig of 1045 bp in length
* 13388 13488: gap of unknown length
* 13488 14638: contig of 1150 bp in length
* 14638 14738: gap of unknown length
* 14738 15895: contig of 1158 bp in length
* 15896 17346: gap of unknown length
* 17346 17347: contig of 1351 bp in length
* 17347 17447: gap of unknown length
* 17447 18910: contig of 1464 bp in length
* 18910 19010: gap of unknown length
* 19011 20474: contig of 1464 bp in length
* 20475 20574: gap of unknown length
* 20575 21945: contig of 1371 bp in length
* 21946 22046: gap of unknown length
* 22046 23658: contig of 1613 bp in length
* 23658 23758: gap of unknown length
* 23758 25591: contig of 1833 bp in length
* 25591 25691: gap of unknown length
* 25692 27779: contig of 2088 bp in length
* 27780 27879: gap of unknown length
* 27880 29540: contig of 1661 bp in length
* 29541 29641: gap of unknown length
* 29641 31165: contig of 1425 bp in length
* 31165: gap of unknown length

* 31166 32789: contig of 1624 bp in length
* 32790 32889: gap of unknown length
* 32890 34164: contig of 1275 bp in length
* 34165 34264: gap of unknown length
* 34265 35468: contig of 1204 bp in length
* 35469 35569: gap of unknown length
* 35569 37527: contig of 1959 bp in length
* 37528 37627: gap of unknown length
* 37628 38999: contig of 2272 bp in length
* 39000 39999: gap of unknown length
* 39999 41707: contig of 1708 bp in length
* 41708 41807: gap of unknown length
* 41808 43703: contig of 1896 bp in length
* 43704 43803: gap of unknown length
* 43804 46114: contig of 2311 bp in length
* 46115 46214: gap of unknown length
* 46215 47315: contig of 1101 bp in length
* 47316 47415: gap of unknown length
* 47416 49427: contig of 2012 bp in length
* 49428 49527: gap of unknown length
* 49528 50778: contig of 1251 bp in length
* 50779 50878: gap of unknown length
* 50879 52271: contig of 1393 bp in length
* 52272 53847: contig of 1476 bp in length
* 53848 53947: gap of unknown length
* 53948 55772: contig of 1825 bp in length
* 55773 55872: gap of unknown length
* 55873 57641: gap of unknown length
* 57642 59120: contig of 1479 bp in length
* 59121 59220: gap of unknown length
* 59221 61342: contig of 2122 bp in length
* 61343 64613: contig of 3171 bp in length
* 64614 64713: gap of unknown length
* 64714 66856: contig of 2143 bp in length
* 66857 66956: gap of unknown length
* 66957 69506: contig of 2550 bp in length
* 69507 69606: gap of unknown length
* 69607 71377: contig of 1771 bp in length
* 71378 71477: gap of unknown length
* 71478 74173: contig of 2696 bp in length
* 74174 74273: gap of unknown length
* 74274 76753: contig of 2480 bp in length
* 76754 76853: gap of unknown length
* 76854 79743: contig of 2790 bp in length
* 79744 81430: contig of 1687 bp in length
* 81431 81530: gap of unknown length
* 81531 84359: contig of 2829 bp in length
* 84360 84459: gap of unknown length
* 84460 86214: contig of 1755 bp in length
* 86215 86315: contig of 3276 bp in length
* 86315 89590: contig of 3276 bp in length

Query Match 80.9%; Score 17.8; DB 2; Length 192579;
Best Local Similarity 90.5%; Pred. No. 1.5e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TGGAGCGCCCGCTGTACTCC 22
|||||
Db 39236 TGGAGCGCTCCATTGTACTCC 39216
|||||

RESULT 14
BD182190/c
LOCUS
DEFINITION
ACCESSION
VERSION

BD182190 1935 bp DNA linear PAT 15-MAY-2003
Polynucleotide probe and primer for detecting beer-clouding lactic
acid bacterium and method of detecting beer-clouding lactic acid
bacterium.
BD182190
BD182190.1 GI:30793108

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 11:31:29 ; Search time 36.7148 Seconds

(without alignments)
2545.580 Million cell updates/sec

Title: US-10-051-769-5

Perfect score: 22

Sequence: 1 aagtgaggttcgaggagctgc 22

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq 29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	22	7	ABX95682 Human Gli
2	22	100.0	22	9	AAD61562 Human Gli
3	22	100.0	261	7	ABX95679 Human Gli
4	22	100.0	261	9	AAD61560 Human Gli
5	22	100.0	263	7	ABX95681 Rat GliTE
6	22	100.0	263	9	AAD61561 Human Gli
7	22	100.0	873	7	ABX95680 Human Gli
8	22	100.0	960	7	ABX95678 Human Gli
9	22	100.0	960	9	AAD61559 Human Gli
10	22	100.0	1092	9	ADE07789 Novel cod
11	22	100.0	2639	9	ADB62530 Human CDN
12	22	100.0	3832	9	AAD61564 Human Gli
13	22	100.0	4450	4	AAK51608 Human pol
14	22	100.0	4470	4	ABA09174 Human sec
15	22	100.0	4470	4	AAK52592 Human pol
16	22	100.0	4470	9	ADE09891 Novel DNA
17	22	100.0	4801	9	ADD49052 Human NOV
18	19.4	88.2	1732	6	ABQ16934 Oligonuc
19	19.4	88.2	1732	6	ABQ16935 Oligonuc
20	18.8	85.5	2249	4	ABL13453 Drosophil
21	18.8	85.5	3582	4	ABL11714 Drosophil
22	18.8	85.5	4382	4	ABL13452 Drosophil
23	18.8	85.5	5973	6	ABS78752 DNA encod

ALIGNMENTS

RESULT 1

ABX95682
ID ABX95682 standard; DNA; 22 BP.
XX
AC ABX95682;
XX
DT 30-JUN-2003 (first entry)
XX
DE Human GliTEN PCR primer #1.
XX
KW Human; ss; PCR; chromosome 10q25; glioblastoma multiforme; brain cancer;
KW GliTEN; GMB; primer.
XX
OS Homo sapiens.
XX
PN US2003044811-A1.
XX
PD 06-MAR-2003.
XX
PF 20-OCT-2001; 2001US-00051769.
XX
PR 20-OCT-2000; 2000US-0242160P.
XX
PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX
PI McKinnon RD;
XX
DR WPI; 2003-418995/39.
XX
PT New isolated nucleic acid representing a gene product associated with glioblastoma multiforme, designated as GliTEN, useful as a molecular marker for tumor cell identification and classification.
XX
PS Claim 5; Page 2; 1lpp; English.
XX
CC The invention relates to an isolated nucleic acid representing a gene product associated with Glioblastoma Multiforme, designated as GliTEN, comprising the human EST (expressed sequence tag) appearing as ABX95678, or a sequence that hybridises under stringent conditions to the EST, or its complement. Also included are a probe for use in identifying a patient at risk for progression into the malignant phenotype, comprising the nucleic acids detailed above, detecting whether a patient is at risk for progression into Glioblastoma Multiforme (comprising: (a) providing a sample from a patient; (b) adding the probe to the sample or performing PCR analysis using the primers appearing as ABX95682 and ABX95683; (c) analysing levels of mRNA bound with the probe; and (d) treating a control

24 18.8 85.5 5202 7 ABX70893
25 18.8 85.5 7584 7 ABZ36402
26 18.8 85.5 26781 4 ABL01850
27 18.8 85.5 35498 4 ABL20316
28 18.4 83.6 2459 4 ABL12495
29 18.4 83.6 6942 4 ABL12494
30 17.8 80.9 190 4 AAI27112
31 17.8 80.9 190 4 ABA75387
32 17.8 80.9 190 4 AAI55954
33 17.8 80.9 190 4 ABA40026
34 17.8 80.9 190 4 AAK50017
35 17.8 80.9 190 4 AAK23948
36 17.8 80.9 190 4 ABS49663
37 17.8 80.9 190 6 ABS23510
38 17.8 80.9 327 8 ACL16432
39 17.8 80.9 336 8 ACL16445
40 17.8 80.9 366 8 ACL16430
41 17.8 80.9 385 8 ACL16414
42 17.8 80.9 426 4 ABL27111
43 17.8 80.9 450 8 ACL16416
44 17.8 80.9 450 8 ACL16447
45 17.8 80.9 460 8 ACL16443

Abx70893 Novel hum
Abz36402 Human sec
Abl01850 Drosophil
Abl20316 Drosophil
Abl12495 Drosophil
Abl12494 Drosophil
Aai27112 Probe #17
Aba75387 Human foe
Aai55954 Probe #24
Aba40026 Probe #18
Aak50017 Human bon
Aak23948 Human bra
Abs49663 Human liv
Abs23510 Human gen
Acl16432 DNA clone
Acl16445 DNA clone
Acl16430 DNA clone
Acl16414 DNA clone
Abl27111 Drosophil
Acl16416 DNA clone
Acl16447 DNA clone
Acl16443 DNA clone

CC sample to assess the level of mRNA in a control sample, where the
 CC presence of increased levels of mRNA expression in the sample in an
 CC amount higher than the control sample indicates risk for progression into
 CC Glioblastoma Multiforme), and a kit for use in detecting whether a
 CC patient is at risk for progression into Glioblastoma Multiforme
 CC (comprising the probe or primers). The probe or the kit is useful for
 CC detecting whether a patient is at risk for progression into Glioblastoma
 CC Multiforme (a type of brain cancer). The EST useful as molecular marker
 CC for tumour cell identification and classification and for diagnosing or
 CC identifying candidates at risk for progression into a malignant phenotype
 CC especially in brain cancer therapy, treatment and diagnosis. The present
 CC sequence is a human GliTEN gene PCR primer
 XX
 SQ Sequence 22 BP; 5 A; 3 C; 10 G; 4 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 7; Length 22;
 Best Local Similarity 100.0%; Pred. No. 4.8;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22

DB 1 AAGGTGGAGTTCGAGGAGCTGC 22

RESULT 2

AAD61562

ID AAD61562 standard; DNA; 22 BP.

XX

AC AAD61562;

XX

DT 15-JAN-2004 (first entry)

XX

DE Human GliTEN cDNA amplifying PCR primer #1.

XX

KW Human; glioblastoma multiforme; GBM; GliTEN; brain cancer; diagnosis;

KW

gene therapy; PCR; primer; ss.

XX

OS Homo sapiens.

XX

PN US2003108915-A1.

XX

PD 12-JUN-2003.

XX

PF 20-AUG-2002; 2002US-00224624.

XX

PR 20-OCT-2000; 2000US-0242160P.

PR

20-OCT-2001; 2001US-00051769.

XX

PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

XX

PI Mckinnon RD;

XX

DR WPI; 2003-810848/76.

XX

PT Novel glioblastoma multiforme associated protein GliTEN useful for

PT

treating glioblastoma multiforme and diagnosing brain cancer.

XX

PS Claim 16; Page 5; 0pp; English.

XX

CC The present invention provides novel glioblastoma multiforme (GBM)
 CC associated protein GliTEN useful for treating glioblastoma multiforme and
 CC diagnosing brain cancer. The invention is useful in brain cancer therapy,
 CC treatment and diagnosis. The invention is also useful in gene therapy.

CC

XX The present sequence is human GliTEN cDNA amplifying PCR primer

XX

SQ Sequence 22 BP; 5 A; 3 C; 10 G; 4 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 9; Length 22;
 Best Local Similarity 100.0%; Pred. No. 4.8;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22

DB 1 AAGGTGGAGTTCGAGGAGCTGC 22

DB 1 AAGGTGGAGTTCGAGGAGCTGC 22

RESULT 3

ABX95679

ID ABX95679 standard; cDNA; 261 BP.

XX

AC ABX95679;

XX

DT 30-JUN-2003 (first entry)

XX

DE Human GliTEN expressed sequence tag.

XX

KW Human; ss; EST; chromosome 10q25; glioblastoma multiforme; brain cancer;
 KW GliTEN; GMB; expressed sequence tag.

XX

OS Homo sapiens.

XX

PN US2003044811-A1.

XX

PD 06-MAR-2003.

XX

PF 20-OCT-2001; 2001US-00051769.

XX

PR 20-OCT-2000; 2000US-0242160P.

XX

PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

XX

PI Mckinnon RD;

XX

DR WPI; 2003-418995/39.

XX

PT New isolated nucleic acid representing a gene product associated with
 PT Glioblastoma Multiforme, designated as GliTEN, useful as a molecular
 PT marker for tumor cell identification and classification.

XX

PS Claim 1; Page 7; 11pp; English.

XX

CC The invention relates to an isolated nucleic acid representing a gene
 CC product associated with Glioblastoma Multiforme, designated as GliTEN,
 CC comprising the human EST (expressed sequence tag) appearing as ABX95678,
 CC or a sequence that hybridises under stringent conditions to the EST, or
 CC its complement. Also included are a probe for use in identifying a
 CC patient at risk for progression into the malignant phenotype, comprising
 CC the nucleic acids detailed above, detecting whether a patient is at risk
 CC for progression into Glioblastoma Multiforme (comprising: (a) providing a
 CC sample from a patient; (b) adding the probe to the sample or performing a
 CC PCR analysis using the primers appearing as ABX95682 and ABX95683; (c)
 CC analysing levels of mRNA bound with the probe; and (d) treating a control
 CC sample to assess the level of mRNA in a control sample, where the
 CC presence of increased levels of mRNA expression in the sample in an
 CC amount higher than the control sample indicates risk for progression into
 CC Glioblastoma Multiforme), and a kit for use in detecting whether a
 CC patient is at risk for progression into Glioblastoma Multiforme
 CC (comprising the probe or primers). The probe or the kit is useful for
 CC detecting whether a patient is at risk for progression into Glioblastoma
 CC Multiforme (a type of brain cancer). The EST useful as molecular marker
 CC for tumour cell identification and classification and for diagnosing or
 CC identifying candidates at risk for progression into a malignant phenotype
 CC especially in brain cancer therapy, treatment and diagnosis. The present
 CC sequence is the human GliTEN expressed sequence tag

SQ Sequence 261 BP; 40 A; 81 C; 102 G; 38 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 7; Length 261;
 Best Local Similarity 100.0%; Pred. No. 5.6;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22

DB 5 AAGGTGGAGTTCGAGGAGCTGC 26

```
RESULT 4
AAD61560
ID AAD61560 standard; cDNA; 261 BP.
XX
AC AAD61560;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human GliTEN EST clone cDNA #1.
XX
KW Human; glioblastoma multiforme; GBM; GliTEN; brain cancer; diagnosis;
KW gene therapy; expressed sequence tag; ss.
XX
OS Homo sapiens.
XX
PN US2003108915-A1.
XX
PD 12-JUN-2003.
XX
PF 20-AUG-2002; 2002US-00224624.
XX
PR 20-OCT-2000; 2000US-0242160P.
XX
PA 20-OCT-2001; 2001US-00051769.
XX
(UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX
PI McKinnon RD;
XX
WPI; 2003-810848/76.
XX
Novel glioblastoma multiforme associated protein GliTEN useful for
PT treating glioblastoma multiforme and diagnosing brain cancer.
XX
Claim 38; Page 11; Opp; English.
XX
The present invention provides novel glioblastoma multiforme (GBM)
CC associated protein GliTEN useful for treating glioblastoma multiforme and
CC diagnosing brain cancer. The invention is useful in brain cancer therapy,
CC treatment and diagnosis. The invention is also useful in gene therapy.
CC The present sequence is human GliTEN EST clone cDNA. This EST clone is
CC used in the exemplification of the invention
XX
SQ Sequence 261 BP; 40 A; 81 C; 102 G; 38 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 9; Length 261;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
Db |||||
5 AAGGTGGAGTTCGAGGAGCTGC 26

RESULT 5
ABX95681
ID ABX95681 standard; cDNA; 263 BP.
XX
AC ABX95681;
XX
DT 30-JUN-2003 (first entry)
XX
DE Rat GliTEN expressed sequence tag.
XX
KW Rat; ss; EST; glioblastoma multiforme; brain cancer; GliTEN; GBM;
KW expressed sequence tag.
XX
OS Rattus Norvegicus.
XX
PN US2003044811-A1.
XX
PD 06-MAR-2003.
XX
PF 20-OCT-2001; 2001US-00051769.

XX
20-OCT-2000; 2000US-0242160P.
XX
(UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX
McKinnon RD;
XX
WPI; 2003-418995/39.
XX
New isolated nucleic acid representing a gene product associated with
PT Glioblastoma Multiforme, designated as GliTEN, useful as a molecular
PT marker for tumor cell identification and classification.
XX
Example 2; Page 4; 11pp; English.
XX
The invention relates to an isolated nucleic acid representing a gene
CC product associated with Glioblastoma Multiforme, designated as GliTEN,
CC comprising the human EST (expressed sequence tag) appearing as ABX95678,
CC or a sequence that hybridises under stringent conditions to the EST, or
CC its complement. Also included are a probe for use in identifying a
CC patient at risk for progression into the malignant phenotype, comprising
CC the nucleic acids detailed above, detecting whether a patient is at risk
CC for progression into Glioblastoma Multiforme (comprising: (a) providing a
CC sample from a patient; (b) adding the probe to the sample or performing a
CC PCR analysis using the primers appearing as ABX95682 and ABX95683; (c)
CC analysing levels of mRNA bound with the probe; and (d) treating a control
CC sample to assess the level of mRNA in a control sample, where the
CC presence of increased levels of mRNA expression in the sample in an
CC amount higher than the control sample indicates risk for progression into
CC Glioblastoma Multiforme), and a kit for use in detecting whether a
CC patient is at risk for progression into Glioblastoma Multiforme
CC (comprising the probe or primers). The probe or the kit is useful for
CC detecting whether a patient is at risk for progression into Glioblastoma
CC Multiforme (a type of brain cancer). The EST useful as molecular marker
CC for tumour cell identification and classification and for diagnosing or
CC identifying candidates at risk for progression into a malignant phenotype
CC especially in brain cancer therapy, treatment and diagnosis. The present
CC sequence is a Rat GliTEN expressed sequence tag
XX
SQ Sequence 263 BP; 41 A; 78 C; 94 G; 50 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 7; Length 263;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
Db |||||
5 AAGGTGGAGTTCGAGGAGCTGC 26

RESULT 6
AAD61561
ID AAD61561 standard; cDNA; 263 BP.
XX
AC AAD61561;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human GliTEN EST clone cDNA #2.
XX
KW Human; glioblastoma multiforme; GBM; GliTEN; brain cancer; diagnosis;
KW gene therapy; expressed sequence tag; ss.
XX
OS Homo sapiens.
XX
PN US2003108915-A1.
XX
PD 12-JUN-2003.
XX
PF 20-AUG-2002; 2002US-00224624.
XX
PR 20-OCT-2000; 2000US-0242160P.
XX
PA 20-OCT-2001; 2001US-00051769.
```

XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX Mckinnon RD;
XX WPI; 2003-810848/76.
XX Novel glioblastoma multiforme associated protein GLITEN useful for
PT treating glioblastoma multiforme and diagnosing brain cancer.
XX Example 2; Page 8; Opp; English.
XX The present invention provides novel glioblastoma multiforme (GBM)
CC associated protein GLITEN useful for treating glioblastoma multiforme and
CC diagnosing brain cancer. The invention is useful in brain cancer therapy,
CC treatment and diagnosis. The invention is also useful in gene therapy.
CC The present sequence is human GLITEN EST clone cDNA. This EST clone is
CC used in the exemplification of the invention
XX
XX Sequence 263 BP; 41 A; 78 C; 94 G; 50 T; 0 U; 0 Other;
Query Match 100.0%; Score 22; DB 9; Length 263;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
Db 5 AAGGTGGAGTTCGAGGAGCTGC 26
RESULT 7
ABX95680
ID ABX95680 standard; DNA; 873 BP.
XX AC ABX95680;
XX 30-JUN-2003 (first entry)
XX Human GLITEN open reading frame.
XX Human; ds; chromosome 10q25; glioblastoma multiforme; brain cancer;
XX GLITEN; GMB.
XX Homo sapiens.
XX Key Location/Qualifiers
FH CDS 1..873
FT /*tag= a
FT /partial
FT /product= "GLITEN"
FT /note= "No stop codon shown"
XX US2003044811-A1.
XX PN 06-MAR-2003.
XX PD 20-OCT-2001; 2001US-00051769.
XX PF 20-OCT-2000; 2000US-0242160P.
XX PR (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX PA Mckinnon RD;
XX PI WPI; 2003-418995/39.
XX New isolated nucleic acid representing a gene product associated with
XX Glioblastoma Multiforme, designated as GLITEN, useful as a molecular
XX marker for tumor cell identification and classification.
XX Disclosure; Page 7; 11pp; English.
XX The invention relates to an isolated nucleic acid representing a gene

CC product associated with Glioblastoma Multiforme, designated as GLITEN,
CC comprising the human EST (expressed sequence tag) appearing as ABX95678,
CC or a sequence that hybridises under stringent conditions to the EST, or
CC its complement. Also included are a probe for use in identifying a
CC patient at risk for progression into the malignant phenotype, comprising
CC the nucleic acids detailed above, detecting whether a patient is at risk
CC for progression into Glioblastoma Multiforme (comprising: (a) providing a
CC sample from a patient; (b) adding the probe to the sample or performing a
CC PCR analysis using the primers appearing as ABX95682 and ABX95683; (c)
CC analysing levels of mRNA bound with the probe; and (d) treating a control
CC sample to assess the level of mRNA in a control sample, where the
CC presence of increased levels of mRNA expression in the sample in an
CC amount higher than the control sample indicates risk for progression into
CC Glioblastoma Multiforme), and a kit for use in detecting whether a
CC patient is at risk for progression into Glioblastoma Multiforme
CC (comprising the probe or primers). The probe or the kit is useful for
CC detecting whether a patient is at risk for progression into Glioblastoma
CC Multiforme (a type of brain cancer). The EST useful as molecular marker
CC for tumour cell identification and classification and for diagnosing or
CC identifying candidates at risk for progression into a malignant phenotype
CC especially in brain cancer therapy, treatment and diagnosis. The present
CC sequence is the open reading frame from the GLITEN gene
XX
XX Sequence 873 BP; 137 A; 304 C; 279 G; 153 T; 0 U; 0 Other;
Query Match 100.0%; Score 22; DB 7; Length 873;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
Db 370 AAGGTGGAGTTCGAGGAGCTGC 391
RESULT 8
ABX95678
ID ABX95678 standard; DNA; 960 BP.
XX AC ABX95678;
XX 30-JUN-2003 (first entry)
XX Human gene encoding GLITEN.
XX Human; ds; gene; chromosome 10q25; glioblastoma multiforme; brain cancer;
XX GLITEN; GMB.
XX Homo sapiens.
XX Key Location/Qualifiers
FH CDS 30..906
FT /*tag= a
FT /product= "GLITEN"
XX US2003044811-A1.
XX PN 06-MAR-2003.
XX PD 20-OCT-2001; 2001US-00051769.
XX PF 20-OCT-2000; 2000US-0242160P.
XX PR (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX PA Mckinnon RD;
XX PI WPI; 2003-418995/39.
XX New isolated nucleic acid representing a gene product associated with
XX Glioblastoma Multiforme, designated as GLITEN, useful as a molecular
XX marker for tumor cell identification and classification.
XX Example 2; Fig 1; 11pp; English.

XX The invention relates to an isolated nucleic acid representing a gene
CC product associated with Glioblastoma Multiforme, designated as GliTEN,
CC comprising the human EST (expressed sequence tag) appearing as ABX95678,
CC or a sequence that hybridises under stringent conditions to the EST, or
CC its complement. Also included are a probe for use in identifying a
CC patient at risk for progression into the malignant phenotype, comprising
CC the nucleic acids detailed above, detecting whether a patient is at risk
CC for progression into Glioblastoma Multiforme (comprising: (a) providing a
CC sample from a patient; (b) adding the probe to the sample or performing
CC PCR analysis using the primers appearing as ABX95682 and ABX95683; (c)
CC analysing levels of mRNA bound with the probe; and (d) treating a control
CC sample to assess the level of mRNA in a control sample, where the
CC presence of increased levels of mRNA expression in the sample in an
CC amount higher than the control sample indicates risk for progression into
CC Glioblastoma Multiforme), and a kit for use in detecting whether a
CC patient is at risk for progression into Glioblastoma Multiforme
CC (comprising the probe or primers). The probe or the kit is useful for
CC detecting whether a patient is at risk for progression into Glioblastoma
CC Multiforme (a type of brain cancer). The EST useful as molecular marker
CC for tumour cell identification and classification and for diagnosing or
CC identifying candidates at risk for progression into a malignant phenotype
CC especially in brain cancer therapy, treatment and diagnosis. The present
CC sequence is a region of human chromosome 10 (q25) containing the GliTEN
CC gene
XX
SQ Sequence 960 BP; 146 A; 324 C; 329 G; 161 T; 0 U; 0 Other;
Query Match 100.0%; Score 22; DB 7; Length 960;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
|||||
Db 400 AAGGTGGAGTTCGAGGAGCTGC 421
RESULT 9
AD61559
ID AAD61559 standard; cDNA; 960 BP.
XX
AC AAD61559;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human GliTEN partial cDNA.
XX
KW Human; glioblastoma multiforme; GBM; GliTEN; brain cancer; diagnosis;
KW gene therapy; chromosome 10; gene; ss.
XX
OS Homo sapiens.
XX
PN US2003108915-A1.
XX
PD 12-JUN-2003.
XX
PF 20-AUG-2002; 2002US-00224624.
XX
PR 20-OCT-2000; 2000US-0242160P.
XX
PR 20-OCT-2001; 2001US-00051769.
XX
PA (UTNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX
PI McKinnon RD;
XX
DR WPI; 2003-810848/76.
XX
PT Novel glioblastoma multiforme associated protein GliTEN useful for
PT treating glioblastoma multiforme and diagnosing brain cancer.
XX
PS Example 2; Fig 1; Opp; English.
XX
SQ The present invention provides novel glioblastoma multiforme (GBM)

CC associated protein GliTEN useful for treating glioblastoma multiforme and
CC diagnosing brain cancer. The invention is useful in brain cancer therapy,
CC treatment and diagnosis. The invention is also useful in gene therapy.
CC The present sequence is human GliTEN partial cDNA. The human GliTEN gene
CC is located on chromosome 10. This partial cDNA is used in the
CC exemplification of the invention
XX
SQ Sequence 960 BP; 146 A; 324 C; 329 G; 161 T; 0 U; 0 Other;
Query Match 100.0%; Score 22; DB 9; Length 960;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
|||||
Db 400 AAGGTGGAGTTCGAGGAGCTGC 421
RESULT 10
ADE07789
ID ADE07789 standard; DNA; 1092 BP.
XX
AC ADE07789;
XX
DT 29-JAN-2004 (first entry)
XX
DE Novel coding sequence (useful for identifying genetic disorders) #855.
XX
KW novel gene; novel protein; tissue marker; molecular weight marker;
KW chromosome marker; genetic disorder; gene; ds.
XX
OS Unidentified.
XX
PN WO2003054152-A2.
XX
PD 03-JUL-2003.
XX
PF 10-DEC-2002; 2002WO-US039555.
XX
PR 10-DEC-2001; 2001US-0339739P.
PR 11-DEC-2001; 2001US-0339453P.
PR 14-MAR-2002; 2002US-0365091P.
PR 14-MAR-2002; 2002US-0365384P.
PR 12-APR-2002; 2002US-0372381P.
PR 12-APR-2002; 2002US-0372615P.
PR 22-APR-2002; 2002US-00128558.
PR 24-APR-2002; 2002US-0376045P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;
PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;
XX
DR WPI; 2003-569235/53.
DR P-PSDB; ADE08700.
XX
PT New polynucleotides, useful for expressing recombinant proteins for
PT analysis, characterization or therapeutic use, or as markers for tissues
PT in which the corresponding protein is preferentially expressed.
XX
PS Claim 1; SEQ ID NO 855; 1177pp; English.
XX
CC The invention comprises the amino acid and coding sequences of novel
CC proteins. The DNA and protein sequences of the invention are useful as:
CC markers for tissues in which the corresponding protein is preferentially
CC expressed; as molecular weight markers on gels; as chromosome markers or
CC tags; to identify chromosomes or to map related gene positions; and to
CC compare with endogenous DNA sequences in patients to identify potential
CC genetic disorders. The present DNA sequence represents a gene of the
CC invention.
XX
SQ Sequence 1092 BP; 190 A; 377 C; 322 G; 203 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 9; Length 1092;
Best Local Similarity 100.0%; Pred. No. 6.2; Mismatches 0; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
|||||
Db 370 AAGGTGGAGTTCGAGGAGCTGC 391
|||||

RESULT 11
ADB62530
ID ADB62530 standard; cDNA; 2639 BP.
XX AC ADB62530;
XX DT 04-DEC-2003 (first entry)
XX DE Human cDNA encoding clone HHDPC20088160.
XX KW Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW tissue regeneration; cell regeneration; membrane protein;
KW signal transduction-related protein; transcription-related protein;
KW osteoporosis; neurological disease; cancer; tumour.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT CDS 214..2639
FT /*tag= a
FT /partial
FT /product= "Clone HHDPC20088160 protein"
XX PN EP1308459-A2.
XX PD 07-MAY-2003.
XX PF 28-MAR-2002; 2002EP-00007401.
XX PR 05-NOV-2001; 2001JP-00379298.
XX PR 25-JAN-2002; 2002US-00350978.
XX PA (HELI-) HELIX RES INST.
XX PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX DR WPI; 2003-450961/43.
XX DR P-PSDB; ADB64500.
XX PT New polynucleotides and polypeptides, useful for developing a diagnostic
PT marker or medicines for regulation of their expression and activity, or
PT as targets of gene therapy.
XX PS Claim 1; Page; 22pp; English.
XX CC The invention discloses a polynucleotide comprising a sequence selected
CC from 1970 fully defined nucleotide sequences which encode novel
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC or its partial peptide, an antibody binding to the polypeptide or peptide
CC of the polynucleotide, immunologically assaying the polypeptide or
CC peptide of the polynucleotide by contacting the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesising the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotides and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell

CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease-related proteins and genes
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a cDNA of the invention. Note: Some of the sequence
CC data for this patent is not represented in the printed specification, but
CC is based on sequence information supplied by the European Patent Office.
XX SQ Sequence 2639 BP; 657 A; 702 C; 719 G; 561 T; 0 U; 0 Other;
Query Match 100.0%; Score 22; DB 9; Length 2639;
Best Local Similarity 100.0%; Pred. No. 6.5; Mismatches 0; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
|||||
Db 583 AAGGTGGAGTTCGAGGAGCTGC 604
|||||

RESULT 12
AAD61564
ID AAD61564 standard; cDNA; 3832 BP.
XX AC AAD61564;
XX DT 15-JAN-2004 (first entry)
XX DE Human GliTEN cDNA.
XX KW Human; glioblastoma multiforme; GBM; GliTEN; brain cancer; diagnosis;
KW gene therapy; chromosome 10; gene; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT CDS 178..3642
FT /*tag= a
FT /product= "Human GliTEN protein"
FT /note= "The CDS is specifically claimed in claim 1"
FT primer_bind 1011..1032
FT /*tag= b
FT primer_bind 1405..1429
FT /bound_moiety= "5' primer #843"
FT /*tag= c
FT primer_bind 2083..2108
FT /bound_moiety= "5' primer #1405"
FT /*tag= d
FT primer_bind 2356..2378
FT /bound_moiety= "5' primer #2083"
FT /*tag= e
FT primer_bind 3625..3650
FT /bound_moiety= "5' primer #2355"
FT /*tag= f
FT primer_bind 3625..3650
FT /bound_moiety= "3' primer #3636"
XX PN US2003108915-A1.
XX PD 12-JUN-2003.
XX PF 20-AUG-2002; 2002US-00224624.
XX PR 20-OCT-2000; 2000US-0242160P.
XX PR 20-OCT-2001; 2001US-00051769.
XX PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX PI McKinnon RD;
XX DR WPI; 2003-810848/76.
XX DR P-PSDB; ABW00758.
XX PT Novel glioblastoma multiforme associated protein GliTEN useful for

PT treating glioblastoma multiforme and diagnosing brain cancer.

PS Claim 1; Fig 4; Opp; English.

XX The present invention provides novel glioblastoma multiforme (GBM)

CC associated protein GliTen useful for treating glioblastoma multiforme and

CC diagnosing brain cancer. The invention is useful in brain cancer therapy,

CC treatment and diagnosis. The invention is also useful in gene therapy.

CC The present sequence is human GliTen cDNA. The human GliTen gene is

XX located on chromosome 10

SQ Sequence 3832 BP; 1090 A; 926 C; 957 G; 859 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 9; Length 3832;

Best Local Similarity 100.0%; Pred. No. 6.7;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22

|||||
Db 547 AAGGTGGAGTTCGAGGAGCTGC 568

RESULT 13

AAK51608

ID AAK51608 standard; cDNA; 4450 BP.

XX AAK51608;

AC AAK51608;

XX 06-NOV-2001 (first entry)

XX Human polynucleotide SEQ ID NO 153.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;

KW nervous system disorder; arthritis; inflammation; ss.

XX Homo sapiens.

OS WO200157190-A2.

PN 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US004098.

XX 03-FEB-2000; 2000US-00496914.

PR 27-APR-2000; 2000US-00560875.

PR 20-JUN-2000; 2000US-00598075.

PR 19-JUL-2000; 2000US-00620325.

PR 01-SEP-2000; 2000US-00654936.

PR 15-SEP-2000; 2000US-00663561.

PR 20-OCT-2000; 2000US-00693325.

PR 30-NOV-2000; 2000US-00728422.

XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;

XX Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WPI; 2001-476283/51.

DR P-PSDB; AM78475.

XX Nucleic acids encoding polypeptides with cytokine-like activities, useful

PT in diagnosis and gene therapy.

PT Claim 1; Page 862-867; 6221pp; English.

PS The invention relates to polynucleotides (AAK51456-AAK53435) and the

XX encoded polypeptides (AAK78323-AAK80302) that exhibit activity relating to

CC cytokine, cell proliferation or cell differentiation or which may induce

CC production of other cytokines in other cell populations. The

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,

CC e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and

CC activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111

CC (AAK52582) and 3666 (AAK80020) are omitted as the relevant pages from the

XX sequence listing were missing at the time of publication

SQ Sequence 4450 BP; 1252 A; 987 C; 1047 G; 1164 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 4; Length 4450;

Best Local Similarity 100.0%; Pred. No. 6.7;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22

|||||
Db 370 AAGGTGGAGTTCGAGGAGCTGC 391

RESULT 14

ABA09174

ID ABA09174 standard; cDNA; 4470 BP.

XX ABA09174;

AC ABA09174;

XX 11-JAN-2002 (first entry)

XX Human secreted protein homologue-encoding cDNA, SEQ ID NO:950.

XX Human; cytokine; cell proliferation; cell differentiation; growth factor;

KW haematopoiesis regulation; tissue growth; immunomodulator; activin;

KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;

KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;

KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;

KW chronic inflammatory condition; proliferative retinopathy;

KW atherosclerosis; coronary heart disease; arterial ischaemia;

KW bone disorder; osteoporosis; vascular growth disorder;

KW tissue regeneration; wound healing; infection; immune disorder;

KW cell culture; drug screening; gene therapy; antiinflammatory;

KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;

KW cytostatic; osteoparatic; vasotropic; cardiant; virucide; antibacterial;

KW antifungal; vulnary; antitumor; ss.

XX Homo sapiens.

OS WO200157188-A2.

PN 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US003800.

PR 03-FEB-2000; 2000US-00496914.

PR 27-APR-2000; 2000US-00560875.

XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-457740/49.

DR P-PSDB; ABE11930.

XX Human proteins and DNA encoding sequences useful for preventing, treating

PT or ameliorating a medical condition in a mammalian subject e.g. arthritis

PT and cancer.

XX Claim 1; Page 813-814; 1963pp; English.

PS Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and

CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The

CC invention also relates to vectors and recombinant host cells comprising a

CC nucleotide of the invention, methods of producing the novel polypeptides,

CC antibodies against the polypeptides, methods of detecting the nucleotides

or polypeptides in a sample, and methods of identifying compounds which bind to polypeptides of the invention. Although novel, many of the polypeptides of the invention have homology to known proteins, thereby giving an insight into their probable biological activities, and hence potential therapeutic applications. The polypeptides of the invention may have various activities, including cytokine, cell proliferation or cell differentiation activities; stem cell growth factor activity; haematopoiesis regulatory activity; tissue growth activity; immunomodulatory activity; activin- or inhibin-related activities; chemotactic or chemokinetic activities; haemostatic, thrombotic or thrombolytic activities; receptor or ligand activities; or may be involved in oncogenesis, cancer cell proliferation or metastasis. Depending on their biological activities, polypeptides and nucleotides of the invention are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Such conditions include cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell disorders), chronic inflammatory conditions (e.g., asthma or arthritis), proliferative retinopathy, atherosclerosis, coronary heart disease, arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal vascular growth. Polypeptides involved with tissue regeneration and repair (or nucleic acids encoding them) may be used to promote wound healing (e.g., of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a cDNA encoding a novel human polypeptide of the invention

Sequence 4470 BP; 1254 A; 991 C; 1054 G; 1171 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 4; Length 4470;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
|||||
Db 370 AAGGTGGAGTTCGAGGAGCTGC 391

RESULT 15

AAKS2592
ID AAKS2592 standard; cDNA; 4470 BP.

AC AAKS2592;

DT 06-NOV-2001 (first entry)

XX Human polynucleotide SEQ ID NO 2121.

DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.

OS Homo sapiens.

XX WO200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US004098.

XX 03-FEB-2000; 2000US-00496914.

PR 27-APR-2000; 2000US-00560875.

PR 20-JUN-2000; 2000US-00598075.

PR 19-JUL-2000; 2000US-00620325.

PR 01-SEP-2000; 2000US-00654936.

PR 15-SEP-2000; 2000US-00663561.

PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;

PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WPI; 2001-476283/51.

DR P-PSDB; AAM79459.

XX Nucleic acids encoding polypeptides with cytokine-like activities, useful in diagnosis and gene therapy.

PS Claim 1; Page 4496-4497; 6221pp; English.

XX The invention relates to polynucleotides (AAKS1456-AAK53435) and the encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the sequence listing were missing at the time of publication

XX Sequence 4470 BP; 1254 A; 991 C; 1054 G; 1171 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 4; Length 4470;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
|||||

Db 370 AAGGTGGAGTTCGAGGAGCTGC 391

Search completed: July 18, 2004, 12:45:04

Job time : 38.7148 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 12:32:20 ; Search time 7.42951 Seconds
(without alignments)
1643.301 Million cell updates/sec

Title: US-10-051-769-5
Perfect score: 22
Sequence: 1 aaggtggagttcgagagctgc 22

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 582709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PTCUS COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	4303	4	US-09-976-594-899
2	17.8	80.9	2184	3	US-09-307-265A-2
3	17.2	78.2	1332	2	US-08-481-814A-3
4	17.2	78.2	1489	3	US-08-836-582-1
5	17.2	78.2	1489	4	US-09-265-566-1
6	17.2	78.2	1489	4	US-09-242-737-3
7	17.2	78.2	1489	4	US-09-023-655-871
8	17.2	78.2	2886	4	US-09-221-017B-863
9	17.2	78.2	3924	3	US-08-726-214-9
10	16.8	76.4	336	4	US-03-821-167-5
11	16.8	76.4	735	3	US-08-998-416-880
12	16.8	76.4	1322	4	US-08-973-005A-1
13	16.8	76.4	1347	4	US-09-016-434-1261
14	16.8	76.4	1650	2	US-08-436-771-10
15	16.8	76.4	1650	2	US-08-434-998-10
16	16.8	76.4	1650	2	US-08-487-797-10
17	16.8	76.4	1650	5	PCT-US95-02058-10
18	16.8	76.4	2213	4	US-09-620-312D-1038
19	16.8	76.4	2214	3	US-08-864-038A-1
20	16.8	76.4	2374	3	US-09-347-801-3
21	16.8	76.4	3331	3	US-08-864-038A-2
22	16.8	76.4	3331	3	US-08-864-038A-4
23	16.8	76.4	5325	4	US-10-164-595-17
24	16.8	76.4	5334	4	US-10-164-595-21
25	16.8	76.4	5581	4	US-10-164-595-19
26	16.8	76.4	11188	4	US-09-821-167-1
27	16.4	74.5	2020	4	US-09-620-312D-751

28	16.4	74.5	3252	2	US-08-770-301A-4	Sequence 4, Appli
29	16.4	74.5	3252	3	US-09-175-581-4	Sequence 4, Appli
30	16.2	73.6	483	4	US-09-252-991A-9327	Sequence 9327, Ap
c 31	16.2	73.6	496	4	US-09-621-976-1717	Sequence 1717, Ap
32	16.2	73.6	531	4	US-09-252-991A-8719	Sequence 8719, Ap
33	16.2	73.6	636	4	US-09-252-991A-8938	Sequence 8938, Ap
34	16.2	73.6	756	4	US-09-252-991A-9832	Sequence 9832, Ap
35	16.2	73.6	762	4	US-09-252-991A-9924	Sequence 9924, Ap
c 36	16.2	73.6	812	3	US-08-998-416-410	Sequence 410, App
c 37	16.2	73.6	1119	4	US-09-252-991A-9448	Sequence 9448, Ap
38	16.2	73.6	1239	4	US-09-252-991A-9383	Sequence 9383, Ap
39	16.2	73.6	1385	4	US-08-932-411A-17	Sequence 17, Appl
40	16.2	73.6	1431	4	US-09-465-558-67	Sequence 67, Appl
41	16.2	73.6	1440	4	US-09-252-991A-12120	Sequence 12120, A
c 42	16.2	73.6	1570	4	US-09-787-583-1	Sequence 1, Appli
43	16.2	73.6	1607	4	US-09-374-454-5	Sequence 5, Appli
44	16.2	73.6	1608	4	US-09-252-991A-11665	Sequence 11665, A
c 45	16.2	73.6	1608	4	US-09-252-991A-11769	Sequence 11769, A

ALIGNMENTS

RESULT 1
US-09-976-594-899
; Sequence 899, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 899
; LENGTH: 4303
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673549 225420.1
US-09-976-594-899

Query Match 100.0%; Score 22; DB 4; Length 4303;
Best Local Similarity 100.0%; Pred. No. 0.83;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGTGGAGTTCGAGGAGCTGC 22
|||||
Db 83 AAGTGGAGTTCGAGGAGCTGC 104

RESULT 2
US-09-307-265A-2
; Sequence 2, Application US/09307265A
; Patent No. 6225456
; GENERAL INFORMATION:
; APPLICANT: Gu, Trent
; APPLICANT: Orita, Satoshi
; APPLICANT: Han, Min
; TITLE OF INVENTION: RAS SUPPRESSOR SUR-5
; FILE REFERENCE: UTC-03732
; CURRENT APPLICATION NUMBER: US/09/307,265A
; CURRENT FILING DATE: 1999-05-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2184
; TYPE: DNA

```
; ORGANISM: Homo sapiens
US-09-307-265A-2

Query Match      80.9%; Score 17.8; DB 3; Length 2184;
Best Local Similarity 90.5%; Pred. No. 53;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGGTGGAGTTCGAGGAGCTGC 22
   |||||||
Db 797 AGCTGGAGTTCGAGGAGCTGC 817

RESULT 3
US-08-481-814A-3
; Sequence 3, Application US/08481814A
; Patent No. 5869040
; GENERAL INFORMATION:
; APPLICANT: Hsu, Yen-Ming
; TITLE OF INVENTION: GENE THERAPY METHODS AND COMPOSITIONS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bogen, Inc.
; STREET: 14 Cambridge Center
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 02142
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481,814A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Kaplan, Warren A
; REFERENCE/DOCKET NUMBER: A001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-679-2838
; TELEFAX: 617-679-2000
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1332 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 63..1301
; OTHER INFORMATION: /function= "transcription factor"
; OTHER INFORMATION: /product= "E2F-4"
US-08-481-814A-3

Query Match      78.2%; Score 17.2; DB 2; Length 1332;
Best Local Similarity 86.4%; Pred. No. 92;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
   |||||||
Db 363 AAGCGAGAGATCGAGGAGCTGC 384

RESULT 4
US-08-836-582-1
; Sequence 1, Application US/08836582
; Patent No. 6045999
; GENERAL INFORMATION:
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Transcription factor E2F-4
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye
; STREET: 1100 No. 630335th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: US
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,582
; FILING DATE: 14-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/00868
; FILING DATE: 18-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9423049.7
; FILING DATE: 15-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Arthur R. Crawford
; REGISTRATION NUMBER: 25,327
; REFERENCE/DOCKET NUMBER: 620-18
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 27..1268
US-08-836-582-1

Query Match      78.2%; Score 17.2; DB 3; Length 1489;
Best Local Similarity 86.4%; Pred. No. 93;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
   |||||||
Db 327 AAGCGAGAGATCGAGGAGCTGC 348

RESULT 5
US-09-265-566-1
; Sequence 1, Application US/09265566
; Patent No. 6303335
; GENERAL INFORMATION:
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Transcription factor E2F-4
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye
; STREET: 1100 No. 630335th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: US
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/09/265,566
FILING DATE: 10-Mar-1999
CLASSIFICATION: <Unknown>
15-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836,582
FILING DATE: <Unknown>
APPLICATION NUMBER: GB 9423049.7
FILING DATE: 15-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Arthur R. Crawford
REGISTRATION NUMBER: 25,327
REFERENCE/DOCKET NUMBER: 620-18
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 27..1268
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-265-566-1

Query Match 78.2%; Score 17.2; DB 4; Length 1489;
Best Local Similarity 86.4%; Pred. No. 93;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAGTGGAGTTCGAGGAGCTGC 22
||||| ||| ||||| |||||
Db 327 AAGCGAGATCGAGGAGCTGC 348

RESULT 6
US-09-242-737-3
Sequence 3, Application US/09242737
Patent No. 6368809
GENERAL INFORMATION:
APPLICANT: BERNARDS, REN
TITLE OF INVENTION: E2F UBIQUITINATION DOMAIN, AND ASSAYS FOR
INHIBITORS OF E2F UBIQUITINATION
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/242,737
FILING DATE: 23-Feb-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB97/02293
FILING DATE: 22-AUG-1997
APPLICATION NUMBER: GB 9617697.9
FILING DATE: 23-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: WILSON, MARY J.
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 620-63
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:

LENGTH: 1489 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 27..1268
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-242-737-3

Query Match 78.2%; Score 17.2; DB 4; Length 1489;
Best Local Similarity 86.4%; Pred. No. 93;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAGTGGAGTTCGAGGAGCTGC 22
||||| ||| ||||| |||||
Db 327 AAGCGAGATCGAGGAGCTGC 348

RESULT 7
US-09-023-655-871
Sequence 871, Application US/09023655
Patent No. 6607879
GENERAL INFORMATION:
APPLICANT: Cocks, Benjamin G.
APPLICANT: Susan G. Stuart
APPLICANT: Jeffrey J. Seilhamer
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
TITLE OF INVENTION: EXPRESSION
NUMBER OF SEQUENCES: 1508
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,655
FILING DATE: HEREWITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0001 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 871:
SEQUENCE CHARACTERISTICS:
LENGTH: 1489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: GI061145
US-09-023-655-871
Query Match 78.2%; Score 17.2; DB 4; Length 1489;
Best Local Similarity 86.4%; Pred. No. 93;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAGTGGAGTTCGAGGAGCTGC 22

; LIFE: DNA
; ORGANISM: Micromonospora rosaria

US-09-821-167-5

Query Match 76.4%; Score 16.8; DB 4; Length 336;
Best Local Similarity 90.0%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGTGGAGTTCGAGGAGCTG 21
|||||
Db 209 AGTGGAGTTCGAGAACCTG 228

RESULT 11

US-08-998-416-880
; Sequence 880, Application US/08998416
; Patent No. 6239264

; GENERAL INFORMATION:

; APPLICANT: Philippsen, Peter

; APPLICANT: Pohlmann, Rainer

; APPLICANT: Steiner, Sabine

; APPLICANT: Mohr, Christine

; APPLICANT: Wendland, Jurgen

; APPLICANT: Knechtle, Philipp

; APPLICANT: Reischung, Corinne

; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPYII
; TITLE OF INVENTION: AND USES THEREOF

; NUMBER OF SEQUENCES: 1152

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 6239264artis Corporation

; STREET: 3054 Cornwallis Road

; CITY: Research Triangle Park

; STATE: No. 6239264th Carolina

; COUNTRY: USA

; ZIP: 27709

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/998,416

; FILING DATE: 24-DEC-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: CH 0016/97

; FILING DATE: 31-DEC-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Meigs, J. Timothy

; REGISTRATION NUMBER: 38,241

; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 919-541-8587

; TELEFAX: 919-541-8689

; INFORMATION FOR SEQ ID NO: 880:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 735 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; ORIGINAL SOURCE:

; ORGANISM: PAG1551UP

US-08-998-416-880

Query Match 76.4%; Score 16.8; DB 3; Length 735;
Best Local Similarity 90.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGTGGAGTTCGAGGAGCTG 21

|||||
Db 114 AGTGGAGTTCGAGGAGCTG 133

RESULT 12

US-08-973-005A-1

; Sequence 1, Application US/08973005A

; Patent No. 6320099

; GENERAL INFORMATION:

; APPLICANT: OGAWA Toshiya

; YOSHIOKA Masaharu

; ISHIDA Isao

; TITLE OF INVENTION: VIRUS RESISTANT PLANTS EXPRESSING ANIMAL

; CELL-DERIVED (2'-5') OLIGOADENYLATE SYNTHETASE AND

; RIBONUCLEASE L AND A METHOD FOR CREATING THE SAME

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W., Suite 500

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/973,005A

; FILING DATE: 01-Dec-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 133741/1995

; FILING DATE: 31-MAY-1995

; APPLICATION NUMBER: JP 285401/1995

; FILING DATE: 01-NOV-1995

; APPLICATION NUMBER: JP 52010/1996

; FILING DATE: 08-MAR-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Bent, Stephen A.

; REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 081356/0113

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)672-5300

; TELEFAX: (202)672-5399

; TELEX: 904136

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1322 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; ORIGINAL SOURCE:

; ORGANISM: Human

; SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-08-973-005A-1

Query Match 76.4%; Score 16.8; DB 4; Length 1322;
Best Local Similarity 90.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GGTGGAGTTCGAGGAGCTGC 22

|||||
Db 465 GGTGGAGTTCGAGTGTCTGC 484

RESULT 13

US-09-016-434-1261

; Sequence 1261, Application US/09016434

; Patent No. 6500938

; GENERAL INFORMATION:

; APPLICANT: Janice Au-Young

; APPLICANT: Jeffrey J. Seilhamer

; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING

; TITLE OF INVENTION: PATHWAY GENE EXPRESSION

; NUMBER OF SEQUENCES: 1490

; CORRESPONDENCE ADDRESS:

ADDRESS: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/016,434
FILING DATE: HERewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0002 US
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1261:
SEQUENCE CHARACTERISTICS:
LENGTH: 1347 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: 9220080
US-09-016-434-1261

Query Match 76.4%; Score 16.8; DB 4; Length 1347;
Best Local Similarity 90.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GGTGGAGTTCGAGGAGCTGC 22
|||||
Db 495 GGTGGAGTTCGATGTGCTGC 514

RESULT 14
US-08-436-771-10
Sequence 10, Application US/08436771
Patent No. 5861300
GENERAL INFORMATION:
APPLICANT: Silverman, Robert H.
TITLE OF INVENTION: Antiviral Transgenic Plants, Vectors,
TITLE OF INVENTION: Cells and Methods
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &
ADDRESSEE: Russell
STREET: 200 E. Broward Boulevard
CITY: Fort Lauderdale
STATE: Florida
COUNTRY: USA
ZIP: 33301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/436,771
FILING DATE:
CLASSIFICATION: 536

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/198,973
FILING DATE: 18-FEB-1994
ATTORNEY/AGENT INFORMATION:
NAME: Manso, Peter J.
REGISTRATION NUMBER: 32,264
REFERENCE/DOCKET NUMBER: CL11363-16
TELEPHONE: 305/527/2498
TELEFAX: 305/764/4996
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1650 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-436-771-10

Query Match 76.4%; Score 16.8; DB 2; Length 1650;
Best Local Similarity 90.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GGTGGAGTTCGAGGAGCTGC 22
|||||
Db 526 GGTGGAGTTCGATGTGCTGC 545

RESULT 15
US-08-434-998-10
Sequence 10, Application US/08434998
Patent No. 5866781
GENERAL INFORMATION:
APPLICANT: Silverman, Robert H.
TITLE OF INVENTION: Antiviral Transgenic Plants, Vectors,
TITLE OF INVENTION: Cells and Methods
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &
ADDRESSEE: Russell
STREET: 200 E. Broward Boulevard
CITY: Fort Lauderdale
STATE: Florida
COUNTRY: USA
ZIP: 33301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/434,998
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/198,973
FILING DATE: 18-FEB-1994
ATTORNEY/AGENT INFORMATION:
NAME: Manso, Peter J.
REGISTRATION NUMBER: 32,264
REFERENCE/DOCKET NUMBER: CL11363-16
TELEPHONE: 305/527/2498
TELEFAX: 305/764/4996
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1650 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-434-998-10

Query Match 76.4%; Score 16.8; DB 2; Length 1650;
Best Local Similarity 90.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GGTGGAGTTCGAGGAGCTGC 22
| | | | | | | | | | | | | | | | | | | | | |
Db 526 GGTGGAGTTCGATGTGCTGC 545

Search completed: July 18, 2004, 14:38:29
Job time : 8.42951 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 12:36:30 ; Search time 38.5902 Seconds
(without alignments)
2781.271 Million cell updates/sec

Title: US-10-051-769-5

Perfect score: 22

Sequence: 1 aaggtggaggttcgaggagctgc 22

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3190992 seqs, 2439311697 residues

Total number of hits satisfying chosen parameters: 6381984

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	22	100.0	22	15	US-10-051-769-5
2	22	100.0	22	15	US-10-224-624-5
3	22	100.0	261	15	US-10-051-769-2
4	22	100.0	261	15	US-10-224-624-2
5	22	100.0	263	15	US-10-051-769-4
6	22	100.0	263	15	US-10-224-624-4
7	22	100.0	873	15	US-10-051-769-3
8	22	100.0	873	15	US-10-224-624-3
9	22	100.0	960	15	US-10-051-769-1
10	22	100.0	960	15	US-10-224-624-1
11	22	100.0	2639	16	US-10-104-047-684
12	22	100.0	3465	15	US-10-224-624-9
13	22	100.0	3649	13	US-10-112-944-63
14	22	100.0	3832	15	US-10-224-624-7

15	22	100.0	4470	13	US-10-276-774-950
16	22	100.0	4801	12	US-10-336-603A-25
17	18.8	85.5	5973	16	US-10-094-466-63
18	18.8	85.5	6202	16	US-10-120-988-120
19	18.4	83.6	844	13	US-10-424-599-26004
20	17.8	80.9	190	9	US-09-864-761-25346
21	17.8	80.9	576	9	US-09-864-761-8627
22	17.8	80.9	762	13	US-10-425-114-23935
23	17.8	80.9	775	13	US-10-425-114-28558
24	17.8	80.9	881	17	US-10-437-963-41316
25	17.8	80.9	1335	15	US-10-228-063-21
26	17.8	80.9	1434	15	US-10-228-063-39
27	17.8	80.9	1436	15	US-10-228-063-43
28	17.8	80.9	1530	16	US-10-369-493-39045
29	17.8	80.9	1530	16	US-10-369-493-39412
30	17.8	80.9	1530	16	US-10-369-493-39797
31	17.8	80.9	1914	16	US-10-094-749-1451
32	17.8	80.9	2019	9	US-09-882-836-3
33	17.8	80.9	2019	13	US-10-377-097-153
34	17.8	80.9	2019	13	US-10-184-648-79
35	17.8	80.9	2477	13	US-09-823-245A-352
36	17.8	80.9	2625	9	US-09-882-836-1
37	17.8	80.9	2625	13	US-10-377-097-151
38	17.8	80.9	2625	13	US-10-184-648-77
39	17.8	80.9	6094	13	US-10-262-839-145
40	17.8	80.9	6855	17	US-10-188-832-65
41	17.8	80.9	233060	13	US-10-087-192-97
42	17.4	79.1	793	17	US-10-437-963-87995
43	17.4	79.1	1272	16	US-10-369-493-43213
44	17.4	79.1	2670	16	US-10-369-493-32390
45	17.2	78.2	152	9	US-09-923-876-3957

ALIGNMENTS

RESULT 1
US-10-051-769-5
; Sequence 5, Application US/10051769
; Publication No. US20030044811A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randy D.
; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION
; FILE REFERENCE: 268/260 (RWJ-00-37)
; CURRENT APPLICATION NUMBER: US/10/051.769
; CURRENT FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,160
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 22
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Primer
US-10-051-769-5

Query Match 100.0%; Score 22; DB 15; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGTGGAGTTCGAGGAGCTGC 22
|||||
Db 1 AAGTGGAGTTCGAGGAGCTGC 22

RESULT 2
US-10-224-624-5
; Sequence 5, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randall D.

```
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GliTEN
; FILE REFERENCE: 54704.8059.US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 261
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Primer
US-10-224-624-5

Query Match      100.0%; Score 22; DB 15; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AAGGTGGAGTTCGAGGAGCTGC 22
      |||
Db      1 AAGGTGGAGTTCGAGGAGCTGC 26

RESULT 3
US-10-051-769-2
; Sequence 2, Application US/10051769
; Publication No. US20030044811A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randy D.
; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION
; FILE REFERENCE: 268/260 (RWJ-00-37)
; CURRENT APPLICATION NUMBER: US/10/051,769
; CURRENT FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 261
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-051-769-2

Query Match      100.0%; Score 22; DB 15; Length 261;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AAGGTGGAGTTCGAGGAGCTGC 22
      |||
Db      1 AAGGTGGAGTTCGAGGAGCTGC 26

RESULT 4
US-10-224-624-2
; Sequence 2, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GliTEN
; FILE REFERENCE: 54704.8059.US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
```

```
; LENGTH: 261
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-224-624-2

Query Match      100.0%; Score 22; DB 15; Length 261;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AAGGTGGAGTTCGAGGAGCTGC 22
      |||
Db      5 AAGGTGGAGTTCGAGGAGCTGC 26

RESULT 5
US-10-051-769-4
; Sequence 4, Application US/10051769
; Publication No. US20030044811A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randy D.
; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION
; FILE REFERENCE: 268/260 (RWJ-00-37)
; CURRENT APPLICATION NUMBER: US/10/051,769
; CURRENT FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 263
; TYPE: DNA
; ORGANISM: Sprague Dawley rat
US-10-051-769-4

Query Match      100.0%; Score 22; DB 15; Length 263;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AAGGTGGAGTTCGAGGAGCTGC 22
      |||
Db      5 AAGGTGGAGTTCGAGGAGCTGC 26

RESULT 6
US-10-224-624-4
; Sequence 4, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GliTEN
; FILE REFERENCE: 54704.8059.US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 263
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-224-624-4

Query Match      100.0%; Score 22; DB 15; Length 263;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AAGGTGGAGTTCGAGGAGCTGC 22
      |||
Db      5 AAGGTGGAGTTCGAGGAGCTGC 26
```

```
RESULT 7
US-10-051-769-3
; Sequence 3, Application US/10051769
; Publication No. US2003004481A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randy D.
; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION
; FILE REFERENCE: 268/260 (RWJ-00-37)
; CURRENT APPLICATION NUMBER: US/10/051,769
; CURRENT FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,160
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 873
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-051-769-3

Query Match      100.0%; Score 22; DB 15; Length 873;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AAGGTGGAGTTCGAGGAGCTGC 22
Db      370 AAGGTGGAGTTCGAGGAGCTGC 391

RESULT 8
US-10-224-624-3
; Sequence 3, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GLITEN
; FILE REFERENCE: 54704.8059.US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 873
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-224-624-3

Query Match      100.0%; Score 22; DB 15; Length 873;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AAGGTGGAGTTCGAGGAGCTGC 22
Db      370 AAGGTGGAGTTCGAGGAGCTGC 391

RESULT 9
US-10-051-769-1
; Sequence 1, Application US/10051769
; Publication No. US2003004481A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randy D.
; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION
; FILE REFERENCE: 268/260 (RWJ-00-37)
; CURRENT APPLICATION NUMBER: US/10/051,769
; CURRENT FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,160
; PRIOR FILING DATE: 2000-10-20
```

```
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-051-769-1

Query Match      100.0%; Score 22; DB 15; Length 960;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AAGGTGGAGTTCGAGGAGCTGC 22
Db      400 AAGGTGGAGTTCGAGGAGCTGC 421

RESULT 10
US-10-224-624-1
; Sequence 1, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GLITEN
; FILE REFERENCE: 54704.8059.US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-224-624-1

Query Match      100.0%; Score 22; DB 15; Length 960;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AAGGTGGAGTTCGAGGAGCTGC 22
Db      400 AAGGTGGAGTTCGAGGAGCTGC 421

RESULT 11
US-10-104-047-684
; Sequence 684, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 684
; LENGTH: 2639
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-684

Query Match      100.0%; Score 22; DB 16; Length 2639;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AAGGTGGAGTTCGAGGAGCTGC 22
```

```
|||||
Db 583 AAGGTGGAGTTCGAGGAGCTGC 604

RESULT 12
US-10-224-624-9
; Sequence 9, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: McKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GlITEN
; FILE REFERENCE: 54704.8059-US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 3465
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-224-624-9

Query Match 100.0%; Score 22; DB 15; Length 3465;
Best Local Similarity 100.0%; Pred. No. 2.2; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
|||||
Db 370 AAGGTGGAGTTCGAGGAGCTGC 391

RESULT 13
US-10-112-944-63
; Sequence 63, Application US/10112944
; Publication No. US20040048249A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yang, Yonghong
; APPLICANT: Weng, Gezhi
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Xue, Aidong J.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wehrman, Tom
; APPLICANT: Ghosh, Malabika
; APPLICANT: Wang, Dunrui
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Zhiwei
; TITLE OF INVENTION: No. US20040048249A1el Nucleic Acids and
; FILE REFERENCE: 805A
; CURRENT APPLICATION NUMBER: US/10/112,944
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: US 09/519,705
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: US 09/552,929
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/577,408
; PRIOR FILING DATE: 2000-05-18

; NUMBER OF SEQ ID NOS: 924
; SOFTWARE: pt_FL_genes Version 5.0
; SEQ ID NO 63
; LENGTH: 3649
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3462)
US-10-112-944-63

Query Match 100.0%; Score 22; DB 13; Length 3649;
Best Local Similarity 100.0%; Pred. No. 2.1; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
|||||
Db 370 AAGGTGGAGTTCGAGGAGCTGC 391

RESULT 14
US-10-224-624-7
; Sequence 7, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: McKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GlITEN
; FILE REFERENCE: 54704.8059-US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 3832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (178)..(3639)
; OTHER INFORMATION:
US-10-224-624-7

Query Match 100.0%; Score 22; DB 15; Length 3832;
Best Local Similarity 100.0%; Pred. No. 2.1; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
|||||
Db 547 AAGGTGGAGTTCGAGGAGCTGC 568

RESULT 15
US-10-276-774-950
; Sequence 950, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y. Tom et al
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 950
```



```
; LENGTH: 4470
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-276-774-950

Query Match      100.0%; Score 22; DB 13; Length 4470;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AAGGTGGAGTTCGAGGAGCTGC 22
        |||||||
Db      370 AAGGTGGAGTTCGAGGAGCTGC 391

Search completed: July 18, 2004, 14:47:31
Job time : 38.5902 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 11:55:38 ; Search time 272.872 Seconds
(without alignments)
2407.604 Million cell updates/sec

Title: US-10-051-769-5

Perfect score: 22
Sequence: 1 aaggtagagtcgaggaagctgc 22

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estov:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_tod:*

26: em_gss_pmg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	437	13	BY276266
2	22	100.0	458	10	BB638573
3	22	100.0	583	10	BF330450
4	22	100.0	667	13	BY734796

5	22	100.0	702	13	BU706646
6	22	100.0	723	9	AW013379
7	22	100.0	738	29	CC567140
8	22	100.0	930	13	BQ960960
9	22	100.0	944	13	BX345030
10	22	100.0	2052	11	AK041847
11	22	100.0	3069	11	BC033019
12	20.4	92.7	681	10	BB866050
13	20.4	92.7	730	13	BU449181
14	20.4	92.7	898	13	BU453045
15	20	90.9	989	13	BX341552
16	19.4	88.2	411	10	BE588105
17	19	86.4	610	10	BE469647
18	19	86.4	910	10	BE469591
19	18.8	85.5	345	29	CC536807
20	18.8	85.5	369	12	BI481867
21	18.8	85.5	404	12	BI361904
22	18.8	85.5	412	9	AI519791
23	18.8	85.5	414	10	BF848453
24	18.8	85.5	425	13	BY260653
25	18.8	85.5	426	10	BE484892
26	18.8	85.5	439	28	AZ048370
27	18.8	85.5	445	9	AA820376
28	18.8	85.5	448	9	AA246831
29	18.8	85.5	456	9	AI518512
30	18.8	85.5	462	12	BI365738
31	18.8	85.5	473	12	BI228997
32	18.8	85.5	503	12	BI172696
33	18.8	85.5	504	12	BM256280
34	18.8	85.5	523	13	BU443584
35	18.8	85.5	524	10	BE236368
36	18.8	85.5	532	12	BI236723
37	18.8	85.5	534	9	AI512754
38	18.8	85.5	547	9	AV665832
39	18.8	85.5	559	14	CF365905
40	18.8	85.5	573	12	BM487827
41	18.8	85.5	576	12	BI539978
42	18.8	85.5	598	9	AI511790
43	18.8	85.5	603	12	BI364150
44	18.8	85.5	604	12	BI216164
45	18.8	85.5	608	9	AI516334

ALIGNMENTS

RESULT 1	BY276266	437 bp	mRNA	linear	EST 11-DEC-2002
LOCUS	BY276266	CDNA clone K430318E15 5', mRNA sequence.			
DEFINITION	BY276266	RIKEN full-length enriched, visual cortex Mus musculus			
ACCESSION	BY276266	GI:26466603			
VERSION	BY276266.1	GI:26466603			
KEYWORDS	EST.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	1 (bases 1 to 437)				
AUTHORS	Nikaido, I., Furuno, M., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusci, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierzki, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Favan, W.J., Pertea, G., Pesole, G.,				

Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reid, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavalan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, F., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12456851

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome-res@gscc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Aizawa, K., Akimura, T., Arakawa, F., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, N., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing Pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in RIKEN Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Michela Fagioli and Takao K. Hensch (Laboratory for Neuronal Circuit Development Brain Science Institute RIKEN 2-1 Hirosawa, Wako-shi, Saitama 351-0198 Japan) whose assistance we gratefully acknowledge. Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

source

```
1. .437
/location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K430318E15"
/tissue_type="visual cortex"
/clone_lib="RIKEN full-length enriched, visual cortex"
```

ORIGIN

Query Match 100.0%; Score 22; DB 13; Length 437;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGCTTCAGGAGCTGC 22
|||||
DB 85 AAGGTGGAGCTTCAGGAGCTGC 106
|||||

RESULT 2

BB638573

LOCUS

DEFINITION

musculus cDNA clone A630041P07 5', mRNA sequence.

BB638573

VERSION

BB638573.1

KEYWORDS

Mus musculus (house mouse)

SOURCE

Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae; Mus.

REFERENCE

AUTHORS

1 (bases 1 to 458)

Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toyota, T., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gscc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing Pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Yamanaka, I., Kiyosawa, H., Kondo, S., Saito, T., Shinagawa, A., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., Arakawa, T., Ishii, Y. and Hayashizaki, Y.

Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct. Func. Genomics 2 pre, L72-L86 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

Location/Qualifiers

1. .458

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="A630041P07"

/tissue_type="thymus"

/dev_stage="3 days neonate"

/lab_host="DH10B"

/clone_lib="RIKEN full-length enriched, 3 days neonate thymus"

/note="Site 1: Sall; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken

FEATURES

source

```
1. .458
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="A630041P07"
/tissue_type="thymus"
/dev_stage="3 days neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 3 days neonate thymus"
```

/note="Site 1: Sall; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken

BB638573 458 bp mRNA linear EST 31-AUG-2001
musculus cDNA clone A630041P07 5', mRNA sequence.

BB638573

VERSION

BB638573.1

KEYWORDS

Mus musculus (house mouse)

SOURCE

Mammalia: Eutheria: Chordata: Vertebrata: Euteleostomi;

Mammalia: Eutheria: Rodentia: Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 458)

Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toyota, T., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gscc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing Pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Yamanaka, I., Kiyosawa, H., Kondo, S., Saito, T., Shinagawa, A., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., Arakawa, T., Ishii, Y. and Hayashizaki, Y.

Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct. Func. Genomics 2 pre, L72-L86 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

Location/Qualifiers

1. .458

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="A630041P07"

/tissue_type="thymus"

/dev_stage="3 days neonate"

/lab_host="DH10B"

/clone_lib="RIKEN full-length enriched, 3 days neonate thymus"

/note="Site 1: Sall; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken

Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watanishi, A., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

source

```

1. .667
  Location/Qualifiers
    /organism="Mus musculus"
    /mol_type="mRNA"
    /db_xref="taxon:10090"
    /clone="G830026K01"
    /tissue_type="mammary gland"
    /cell_line="RCB-0526 Jyg-MC(A)"
    /clone_lib="RIKEN full-length enriched, mammary gland
    RCB-0526 Jyg-MC(A) cDNA"
  
```

ORIGIN

```

Query Match      100.0%; Score 22; DB 13; Length 667;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
    |||||
DB 560 AAGGTGGAGTTCGAGGAGCTGC 581
    |||||
  
```

```

RESULT 5
LOCUS BU706646
DEFINITION BU706646 702 bp mRNA linear EST 15-JUL-2003
IMAGE:6409978 5', mRNA sequence.
ACCESSION BU706646
VERSION BU706646.1 GI:23637322
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 702)
JOURNAL NIH-MGC http://mgc.nci.nih.gov/
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
  Location/Qualifiers
    1. .702
    /organism="Mus musculus"
  
```

FEATURES

source

```

1. .702
  Location/Qualifiers
    /organism="Mus musculus"
  
```

```

/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6409978"
/tissue_type="whole brain"
/dev_stages="embryo 12.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH-BMAP_F00"
/note="Organ: Brain; Vector: pYX-Asc; Site_1: EcoR I;
Site_2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is TCAGAGAGCC. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."
  
```

ORIGIN

```

Query Match      100.0%; Score 22; DB 13; Length 702;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
    |||||
DB 314 AAGGTGGAGTTCGAGGAGCTGC 335
    |||||
  
```

RESULT 6

```

LOCUS AW013379
DEFINITION AW013379 723 bp mRNA linear EST 10-SEP-1999
sp042ks Winter flounder spleen Pseudopleuronectes americanus cDNA
clone sp042ks 5' similar to C53B4.4 [Caenorhabditis elegans], mRNA
sequence.
ACCESSION AW013379
VERSION AW013379.1 GI:5862157
KEYWORDS EST.
SOURCE Pseudopleuronectes americanus (winter flounder)
ORGANISM Pseudopleuronectes americanus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
TITLE Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
JOURNAL Pleuronectidae; Pleuronectidae; Pseudopleuronectes.
COMMENT 1 (bases 1 to 723)
Contact: Reith M
Marine Biology
NRC Institute for Marine Biosciences
1411 Oxford St., Halifax, Nova Scotia, B3H 3Z1, Canada
Tel: (902) 426-8276
Fax: (902) 426-9413
Email: michael.reith@nrc.ca
Seq primer: M13 Forward
  Location/Qualifiers
    1. .723
    /organism="Pseudopleuronectes americanus"
    /mol_type="mRNA"
    /db_xref="taxon:8265"
    /clone="sp042ks"
    /sex="female"
    /dev_stage="adult"
    /clone_lib="Winter flounder spleen"
    /note="Organ: spleen"
  
```

ORIGIN

Query Match 100.0%; Score 22; DB 9; Length 723;
 Best Local Similarity 100.0%; Pred. No. 4.6e+02;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
 |||||
 Db 486 AAGGTGGAGTTCGAGGAGCTGC 507

RESULT 7

CC567140 738 bp DNA linear GSS 18-JUN-2003
 LOCUS CH240_441017.TARBAC1382 CHORI-240 Bos taurus genomic clone
 DEFINITION CH240_441017, genomic survey sequence.
 CC567140
 VERSION CC567140.1 GI:31900496
 KEYWORDS GSS.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 REFERENCE 1 (bases 1 to 738)
 AUTHORS Holt R., Stott J., Yang G., Barber S., Smailus D., Prabhu A.-L.,
 Tsai M., Cloutier A., Lee D., Gilm N., Olson T., Mayo M.,
 Butterfield Y., Kirkpatrick R., Liu J., Guin R., Chan A., Chiu R.,
 Mathewson C., Wye N., Masson A., Brown-John M., Jones S.,
 Schein J., Marra M., de Jong P., Keele J. W. and Kappes S. M.
 TITLE Bovine BAC End Sequences from Library CHORI-240, PLATES 393 to 478
 JOURNAL Unpublished (2003)
 COMMENT Other_GSSs: CH240_441017.T7
 Contact: Rob Holt
 Sequencing
 The British Columbia Cancer Agency Genome Science Centre
 600 W. 10th Ave, Vancouver, British Columbia, Canada V5Z 4B6
 Tel: 604-877-6085
 Fax: 604-877-6276
 Email: rholt@bcgsc.ca
 Clones are derived from the bovine BAC library CHORI-240
 (http://www.chori.org/bacpac/bovine240.htm). For BAC library
 availability, please contact Pieter de Jong (pjejong@mail.cho.org).
 Clones may be purchased from BACPAC Resources
 (http://www.chori.org/bacpac/ordering/information.htm). This work
 was undertaken as part of the International Bovine BAC Mapping
 Consortium (IBBMC) by CSIRO Livestock Industries, Australia and the
 British Columbia Genome Sciences Centre, Canada.
 Plate: 441 row: 0 column: 17
 Seq primer: SP6
 Class: BAC ends.

FEATURES

source
 Location/Qualifiers
 1..738
 /organism="Bos taurus"
 /mol_type="genomic DNA"
 /strain="breed: Hereford"
 /db_xref="taxon:9913"
 /clone="CH240_441017"
 /sex="Male"
 /cell_type="Blood"
 /clone_lib="CHORI-240"
 /note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
 Hereford bull Li Domino 99375; CHORI-240 Bovine BAC
 library (Male) produced by Pieter de Jong"

ORIGIN

Query Match 100.0%; Score 22; DB 29; Length 738;
 Best Local Similarity 100.0%; Pred. No. 4.6e+02;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
 |||||
 Db 347 AAGGTGGAGTTCGAGGAGCTGC 368

RESULT 8

BQ960960 930 bp mRNA linear EST 21-AUG-2002
 LOCUS AGENCOURT 8955743 NCI CGAP Mam2 Mus musculus cDNA clone
 DEFINITION IMAGE:6433792 5', mRNA sequence.
 BQ960960
 VERSION BQ960960.1 GI:22376438
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 930)
 AUTHORS NIH-MGC http://mgs.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM13960 row: f column: 17
 High quality sequence stop: 603.
 Location/Qualifiers
 1..930
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N-3"
 /db_xref="taxon:10090"
 /clone="IMAGE:6439792"
 /tissue_type="tumor, biopsy sample"
 /dev_stage="5 months"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Mam2"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

FEATURES

source
 Location/Qualifiers
 1..930
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N-3"
 /db_xref="taxon:10090"
 /clone="IMAGE:6439792"
 /tissue_type="tumor, biopsy sample"
 /dev_stage="5 months"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Mam2"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

ORIGIN

Query Match 100.0%; Score 22; DB 13; Length 930;
 Best Local Similarity 100.0%; Pred. No. 4.9e+02;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
 |||||
 Db 339 AAGGTGGAGTTCGAGGAGCTGC 360

RESULT 9

BX345030 944 bp mRNA linear EST 01-MAY-2003
 LOCUS BX345030 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
 DEFINITION cDNA clone CS00DK002YAL2 5-PRIME, mRNA sequence.
 BX345030
 VERSION BX345030.1 GI:30313335
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 944)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France
 Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 5483.r For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS2BAX16ZB10_AX27ZD11.1&cluster=5483.r.
 Contact : Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS2BAX16ZB10_AX27ZD11.1.
 Location/Qualifiers
 1. 944
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DK002YA12"
 /cell_type="HELA CELLS COT 25-NORMALIZED"
 /cell_line="HELA"
 /clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
 Query Match 100.0%; Score 22; DB 13; Length 944;
 Best Local Similarity 100.0%; Pred. No. 5e+02; Mismatches 0; Indels 0; Gaps 0;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
 |||||
 Db 645 AAGGTGGAGTTCGAGGAGCTGC 666
 |||||

RESULT 10
 AK041847
 LOCUS
 DEFINITION Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
 library, clone:A630041P07 product:hypothetical PDZ domain (also
 known as DHR or GLGF) containing protein, full insert sequence.
 AK041847
 VERSION AK041847.1 GI:26088683
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE 2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159
 REFERENCE 3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913

11076861
 4
 REFERENCE
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 REFERENCE
 AUTHORS 5
 TITLE The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II team.
 JOURNAL Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 NATURE 420, 563-573 (2002)
 REFERENCE
 AUTHORS 6 (bases 1 to 2052)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 COMMENT
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
 Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
 Trust/WRC building Addenbrookes Hospital Cambridge) whose
 assistance we gratefully acknowledge.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/
 Location/Qualifiers
 1. 2052
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM DB:A630041P07"
 /db_xref="MGI:2406859"
 /db_xref="taxon:10090"
 /clone="A630041P07"
 /tissue_type="thymus"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="3 days neonate"
 misc_feature
 1. 2052
 /note="hypothetical PDZ domain (also known as DHR or GLGF)
 containing protein (InterPro|IPR001478, evidence:
 InterPro)"

ORIGIN
 Query Match 100.0%; Score 22; DB 11; Length 2052;
 Best Local Similarity 100.0%; Pred. No. 6.2e+02;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
 |||||
 Db 14 AAGGTGGAGTTCGAGGAGCTGC 35
 |||||

RESULT 11


```

BC033019          3069 bp      mRNA      linear      HTC 19-NOV-2003
LOCUS             Homo sapiens hypothetical protein LOC118987, mRNA (cDNA clone
DEFINITION        IMAGE:4821752), containing frame-shift errors.
ACCESSION         BC033019
VERSION           BC033019.1 GI:21542573
KEYWORDS          HTC.
SOURCE            Homo sapiens (human)
ORGANISM          Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE         1 (bases 1 to 3069)
AUTHORS           Klausner,R.D., Collins,F.S., Wagner,L.H., Derge,J.G.,
                  Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
                  Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
                  Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
                  Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
                  Scheetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S.,
                  Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
                  Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J.,
                  McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
                  Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
                  Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
                  Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
                  Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
                  Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
                  Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
                  Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalins,D.E.,
                  Schnurch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE             Generation and initial analysis of more than 15,000 full-length
                  human and mouse cDNA sequences
JOURNAL           Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE           22388257
PUBMED            12477932
REFERENCE         2 (bases 1 to 3069)
AUTHORS           Strausberg,R.
TITLE             Direct Submission
JOURNAL           Submitted (21-JUN-2002) National Institutes of Health, Mammalian
                  Gene Collection (MGC), Cancer Genomics Office, National Cancer
                  Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
                  USA
REMARK            NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT            Contact: MGC help desk
                  Email: gcapsb@mail.nih.gov
                  Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
                  cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
                  Toshiyuki and Piero Carninci (RIKEN)
                  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                  DNA Sequencing by: Institute for Systems Biology
                  http://www.systemsbio.org
                  contact: amadan@systemsbiology.org
                  Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
                  Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 33 Row: b Column: 4
This clone was selected for full length sequencing because it
passed the following selection criteria: GenomeScan gene prediction
This clone has the following problem: frame shifted.
FEATURES             Location/Qualifiers
     source           1..3069
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4821752"
                     /tissue_type="Testis"
                     /clone_lib="NIH MGC 97"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
ORIGIN

```

```

Query Match       100.0%; Score 22; DB 11; Length 3069;
Best Local Similarity 100.0%; Pred. No. 7e+02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY               1 AAGGTGGAGTTCGAGGAGCTGC 22
                  |||||
DB               614 AAGGTGGAGTTCGAGGAGCTGC 635

RESULT 12
BB866050
LOCUS           BB866050          581 bp      mRNA      linear      EST 09-JUL-2003
DEFINITION     BB866050 RIKEN full-length enriched, CRL-1751 WEHI 164 cDNA Mus
                  musculus cDNA clone G431003009 5', mRNA sequence.
ACCESSION      BB866050
VERSION        BB866050.1 GI:17112260
KEYWORDS       EST.
SOURCE         Mus musculus (house mouse)
ORGANISM       Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 581)
AUTHORS        Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
                  Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Inotani,K.,
                  Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Komoto,H., Kouda,M.,
                  Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
                  Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
                  Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
                  Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
                  Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A.,
                  Muramatsu,M. and Hayashizaki,Y.
TITLE          RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
                  2001)
JOURNAL        Unpublished (2001)
COMMENT        Contact: Yoshihide Hayashizaki
                  Laboratory for Genome Exploration Research Group, RIKEN Genomic
                  Sciences Center (GSC), Yokohama Institute
                  The Institute of Physical and Chemical Research (RIKEN)
                  1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
                  Tel: 81-45-503-9222
                  Fax: 81-45-503-9216
                  Email: genome-resgsc.riken.go.jp/
                  URL: http://genome.gsc.riken.go.jp/
                  Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
                  Itoh,M., Komoto,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
                  Normalization and subtraction of cap-trapper-selected cDNAs to
                  prepare full-length cDNA libraries for rapid discovery of new
                  genes. Genome Res. 10 (10), 1617-1630 (2000)
                  wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
                  Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
                  Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.,
                  and Hayashizaki,Y.
                  RIKEN integrated sequence analysis (RISA) system--384-format
                  sequencing pipeline with 384 multicapillary sequencer. Genome Res.
                  10 (11), 1757-1771 (2000)
                  Komno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
                  Sugahara,Y. and Hayashizaki,Y.
                  Computer-based methods for the mouse full-length cDNA
                  encyclopedia: real-time sequence clustering for construction of a
                  nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
                  Please visit our web site (http://genome.gsc.riken.go.jp) for
                  further details.
                  e mouse tissues.
FEATURES             Location/Qualifiers
     source           1..681
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="BALB/c"
                     /db_xref="taxon:10090"
                     /clone="G431003009"
                     /cell_line="CRL-1751 WEHI 164"
                     /clone_lib="RIKEN full-length enriched, CRL-1751 WEHI 164
                     cDNA"
ORIGIN

```

ORIGIN

Query Match 92.7%; Score 20.4; DB 10; Length 681;
 Best Local Similarity 95.5%; Pred. No. 1.8e+03;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
 DB 563 AAGGTGGAGTTCGAGGAGCTGC 584

RESULT 13
 BU449181 730 bp mRNA linear EST 29-NOV-2002
 LOCUS 603767430F1 CSEQBN13 Gallus gallus cDNA clone CHEST69518 5', mRNA
 DEFINITION sequence.
 ACCESSION BU449181
 VERSION BU449181.1 GI:25938492
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 730)
 AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 TITLE A Comprehensive Collection of Chicken cDNAs
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
 MEDLINE 22335534
 PUBMED 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology
 (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES

source
 Location/Qualifiers
 1..730
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="Layer"
 /db_xref="taxon:9031"
 /clone="CHEST69518"
 /sex="Female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="CSEQBN13"
 /note="Organ: ovary; Vector: pBluescript II KS(+); Site_1:
 EcoRI; Site_2: NotI; This normalized library was
 constructed from 1 million independent clones. cDNA
 synthesis was initiated using an oligo(dT) primer, using
 methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was blunted, ligated to NotI adapters, digested with
 EcoRI, size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."

ORIGIN

Query Match 92.7%; Score 20.4; DB 13; Length 730;
 Best Local Similarity 95.5%; Pred. No. 1.8e+03;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
 DB 176 AAGGTGGAGTTCGAGGAGCTGC 197

RESULT 14
 BU453045
 LOCUS 603219082F1 CSEQBN14 Gallus gallus cDNA clone CHEST210c22 5', mRNA
 DEFINITION sequence.
 ACCESSION BU453045
 VERSION BU453045.1 GI:25942356
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 898)
 AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 TITLE A Comprehensive Collection of Chicken cDNAs
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
 MEDLINE 22335534
 PUBMED 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology
 (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES

source
 Location/Qualifiers
 1..898
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="Layer"
 /db_xref="taxon:9031"
 /clone="CHEST210c22"
 /sex="Female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="CSEQBN14"
 /note="Organ: ovary; Vector: pBluescript II KS(+); Site_1:
 EcoRI; Site_2: NotI; This normalized library was
 constructed from 1 million independent clones. cDNA
 synthesis was initiated using an oligo(dT) primer, using
 methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was blunted, ligated to NotI adapters, digested with
 EcoRI, size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."

ORIGIN

Query Match 92.7%; Score 20.4; DB 13; Length 898;
 Best Local Similarity 95.5%; Pred. No. 1.9e+03;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AAGGTGGAGTTCGAGGAGCTGC 22
 DB 176 AAGGTGGAGTTCGAGGAGCTGC 197

RESULT 15
 BX341552
 LOCUS BX341552 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
 DEFINITION CDNA clone CSODK002YA12 5-PRIME, mRNA sequence.
 ACCESSION BX341552
 VERSION BX341552.1 GI:30339998

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
(bases 1 to 989)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5483.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DK002BA06QPl&cluster=5483.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DK002BA06QPl.
Location/Qualifiers
1. .989
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK002YA12"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Query Match 90.9%; Score 20; DB 13; Length 989;
Best Local Similarity 90.9%; Pred.No. 2.8e+03;
Matches 20; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAGTGGAGTTCGAGGAGCTGC 22
|||
Db 637 AARGTGAATTCGAGGAGCTGC 658
Search completed: July 18, 2004, 14:36:35
Job time : 275.872 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 11:31:29 ; Search time 36.7148 Seconds
(without alignments)
2545.580 Million cell updates/sec

Title: US-10-051-769-6
Perfect score: 22
Sequence: 1 gtggaagccgcgtgtactcc 22

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : N Geneseq 29Jan04:*
- 1: geneseqn1980s:*
 - 2: Geneseqn1990s:*
 - 3: Geneseqn2000s:*
 - 4: Geneseqn2001as:*
 - 5: Geneseqn2001bs:*
 - 6: Geneseqn2002as:*
 - 7: Geneseqn2003as:*
 - 8: Geneseqn2003bs:*
 - 9: Geneseqn2003cs:*
 - 10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	22	7 ABX95683	Abx95683 Human Gli
2	22	100.0	22	9 AAD61563	Aad61563 Human Gli
C 3	22	100.0	263	7 ABX95681	Abx95681 Rat GliTE
C 4	22	100.0	263	9 AAD61561	Aad61561 Human Gli
C 5	20.4	92.7	261	7 ABX95679	Abx95679 Human Gli
C 6	20.4	92.7	261	9 AAD61560	Aad61560 Human Gli
C 7	20.4	92.7	873	7 ABX95680	Abx95680 Human Gli
C 8	20.4	92.7	960	7 ABX95678	Abx95678 Human gen
C 9	20.4	92.7	960	9 AAD61559	Aad61559 Human Gli
C 10	20.4	92.7	1092	9 ADE07789	Ade07789 Novel cod
C 11	20.4	92.7	2639	9 ADB62530	Adb62530 Human cDN
C 12	20.4	92.7	3832	9 AAD61564	Aad61564 Human Gli
C 13	20.4	92.7	3847	6 ABX87764	Abx87764 Human cDN
C 14	20.4	92.7	4450	4 AAK51608	Aak51608 Human pol
C 15	20.4	92.7	4470	4 ABA09174	Aba09174 Human sec
C 16	20.4	92.7	4470	4 AAK52592	Aak52592 Human pol
C 17	20.4	92.7	4470	9 ADE09891	Ade09891 Novel DNA
C 18	20.4	92.7	4801	9 ADD49052	Add49052 Human NOV
C 19	18.8	85.5	879	2 AAQ24436	Aaq24436 Mutant Cy
C 20	18.8	85.5	879	2 AAQ24437	Aaq24437 Mutated m
C 21	18.8	85.5	960	2 AAQ36210	Aaq36210 DNA encod
C 22	17.4	79.1	1935	7 ABZ76503	Abz76503 Lactobaci
C 23	17.4	79.1	8056	7 ABZ76501	Abz76501 Lactobaci

C 24	17.4	79.1	9901	7 ABZ76576	Abz76576 Lactobaci
C 25	17.2	78.2	618	2 AAT46934	Aat46934 Human von
C 26	17.2	78.2	681	2 AAQ13428	Aaq13428 vWF GPIb
C 27	17.2	78.2	681	2 AAQ78689	Aaq78689 vWF GPIb
C 28	17.2	78.2	681	2 AAV08900	Aav08900 Von Wille
C 29	17.2	78.2	681	2 AAV74157	Aav74157 Human mat
C 30	17.2	78.2	681	3 AAZ56182	Aaz56182 Plasmid p
C 31	17.2	78.2	750	2 AAQ46008	Aaq46008 cDNA enco
C 32	17.2	78.2	893	2 AAQ14803	Aaq14803 Von Wille
C 33	17.2	78.2	893	2 AAQ48233	Aaq48233 von Wille
C 34	17.2	78.2	1272	3 AAA15183	Aaa15183 DNA encod
C 35	17.2	78.2	1627	4 AAH14140	Aah14140 Human cDN
C 36	17.2	78.2	2460	7 RAD56131	Rad56131 Human NOT
C 37	17.2	78.2	2460	8 ADA02494	Ada02494 Human NOT
C 38	17.2	78.2	2460	9 ADB72232	Adb72232 Human NOT
C 39	17.2	78.2	2591	2 AAQ45961	Aaq45961 HSA-vWF (4
C 40	17.2	78.2	3234	2 AAQ30999	Aaq30999 Notch c10
C 41	17.2	78.2	3234	2 AAL55428	Aal55428 Human Not
C 42	17.2	78.2	6153	2 AAV08901	Aav08901 Von Wille
C 43	17.2	78.2	6153	2 AAV74156	Aav74156 Human mat
C 44	17.2	78.2	6153	3 AAZ56177	Aaz56177 Human von
C 45	17.2	78.2	6360	6 ABL66946	AbL66946 Lung canc

ALIGNMENTS

RESULT 1
ABX95683
ID ABX95683 standard; DNA; 22 BP.
XX
AC ABX95683;
XX
DT 30-JUN-2003 (first entry)
XX
DE Human GLITEN PCR primer #2.
XX
KW Human; ss; PCR; chromosome 10q25; glioblastoma multiforme; brain cancer;
KW GLITEN; GMB; primer.
XX
OS Homo sapiens.
XX
DN US2003044811-A1.
XX
PD 06-MAR-2003.
XX
PF 20-OCT-2001; 2001US-00051769.
XX
PR 20-OCT-2000; 2000US-0242160P.
XX
FA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX
PI McKinnon RD;
XX
DR WPI; 2003-418995/39.
XX
PT New isolated nucleic acid representing a gene product associated with
PT Glioblastoma Multiforme, designated as GLITEN, useful as a molecular
XX marker for tumor cell identification and classification.
XX
PS Claim 5; Page 2; 11pp; English.
XX
CC The invention relates to an isolated nucleic acid representing a gene
CC product associated with Glioblastoma Multiforme, designated as GLITEN,
CC comprising the human EST (expressed sequence tag) appearing as ABX95678,
CC or a sequence that hybridises under stringent conditions to the EST, or
CC its complement. Also included are a probe for use in identifying a
CC patient at risk for progression into the malignant phenotype, comprising
CC the nucleic acids detailed above, detecting whether a patient is at risk
CC for progression into Glioblastoma Multiforme (comprising: (a) providing a
CC sample from a patient; (b) adding the probe to the sample or performing a
CC PCR analysis using the primers appearing as ABX95682 and ABX95683; (c)
CC analysing levels of mRNA bound with the probe; and (d) treating a control

CC sample to assess the level of mRNA in a control sample, where the
CC presence of increased levels of mRNA expression in the sample in an
CC amount higher than the control sample indicates risk for progression into
CC Glioblastoma Multiforme), and a kit for use in detecting whether a
CC patient is at risk for progression into Glioblastoma Multiforme
CC (comprising the probe or primers). The probe or the kit is useful for
CC detecting whether a patient is at risk for progression into Glioblastoma
CC Multiforme (a type of brain cancer). The EST useful as molecular marker
CC for tumour cell identification and classification and for diagnosing or
CC identifying candidates at risk for progression into a malignant phenotype
CC especially in brain cancer therapy, treatment and diagnosis. The present
CC sequence is a human GliTEN gene PCR primer
XX
SQ Sequence 22 BP; 3 A; 7 C; 7 G; 5 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 7; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.98;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGAAGCCGCCGTTGTACTCC 22
DB 1 GTGGAAGCCGCCGTTGTACTCC 22

RESULT 2
AAD61563
ID AAD61563 standard; DNA; 22 BP.

XX AAD61563;
XX
XX 15-JAN-2004 (first entry)

DE Human GliTEN cDNA amplifying PCR primer #2.

XX Human; glioblastoma multiforme; GBM; GliTEN; brain cancer; diagnosis;
KW gene therapy; PCR; primer; ss.

XX Homo sapiens.

XX US2003108915-A1.

XX 12-JUN-2003.

XX 20-AUG-2002; 2002US-00224624.

XX 20-OCT-2000; 2000US-0242160P.

PR 20-OCT-2001; 2001US-00051769.

XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

XX Mckinnon RD;

XX WPI; 2003-810848/76.

XX Novel glioblastoma multiforme associated protein GliTEN useful for
PT treating glioblastoma multiforme and diagnosing brain cancer.

XX Claim 16; Page 5; Opp; English.

XX The present invention provides novel glioblastoma multiforme (GBM)
CC associated protein GliTEN useful for treating glioblastoma multiforme and
CC diagnosing brain cancer. The invention is useful in brain cancer therapy,
CC treatment and diagnosis. The invention is also useful in gene therapy.
CC The present sequence is human GliTEN cDNA amplifying PCR primer
XX

SQ Sequence 22 BP; 3 A; 7 C; 7 G; 5 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 9; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.98;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGAAGCCGCCGTTGTACTCC 22
DB 1 GTGGAAGCCGCCGTTGTACTCC 22

DB 1 GTGGAAGCCGCCGTTGTACTCC 22

RESULT 3

ABX95681/c

ID ABX95681 standard; cDNA; 263 BP.

XX ABX95681;

XX 30-JUN-2003 (first entry)

XX Rat GliTEN expressed sequence tag.

XX Rat; ss; EST; glioblastoma multiforme; brain cancer; GliTEN; GBM;
XX expressed sequence tag.

XX Rattus Norvegicus.

XX US2003044811-A1.

XX 06-MAR-2003.

XX 20-OCT-2001; 2001US-00051769.

XX 20-OCT-2000; 2000US-0242160P.

XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

XX Mckinnon RD;

XX WPI; 2003-418995/39.

XX New isolated nucleic acid representing a gene product associated with
XX Glioblastoma Multiforme, designated as GliTEN, useful as a molecular
XX marker for tumor cell identification and classification.

XX Example 2; Page 4; lipp; English.

XX The invention relates to an isolated nucleic acid representing a gene
CC product associated with Glioblastoma Multiforme, designated as GliTEN,
CC comprising the human EST (expressed sequence tag) appearing as ABX95678,
CC or a sequence that hybridises under stringent conditions to the EST, or
CC its complement. Also included are a probe for use in identifying a
CC patient at risk for progression into the malignant phenotype, comprising
CC the nucleic acids detailed above, detecting whether a patient is at risk
CC for progression into Glioblastoma Multiforme (comprising: (a) providing a
CC sample from a patient; (b) adding the probe to the sample or performing a
CC PCR analysis using the primers appearing as ABX95682 and ABX95683; (c)
CC analysing levels of mRNA bound with the probe; and (d) treating a control
CC sample to assess the level of mRNA in a control sample, where the
CC presence of increased levels of mRNA expression in the sample in an
CC amount higher than the control sample indicates risk for progression into
CC Glioblastoma Multiforme), and a kit for use in detecting whether a
CC patient is at risk for progression into Glioblastoma Multiforme
CC (comprising the probe or primers). The probe or the kit is useful for
CC detecting whether a patient is at risk for progression into Glioblastoma
CC Multiforme (a type of brain cancer). The EST useful as molecular marker
CC for tumour cell identification and classification and for diagnosing or
CC identifying candidates at risk for progression into a malignant phenotype
CC especially in brain cancer therapy, treatment and diagnosis. The present
CC sequence is a Rat GliTEN expressed sequence tag
XX

SQ Sequence 263 BP; 41 A; 78 C; 94 G; 50 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 7; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGAAGCCGCCGTTGTACTCC 22
DB 244 GTGGAAGCCGCCGTTGTACTCC 223

```
RESULT 4
AAD61561/c
ID AAD61561 standard; cDNA; 263 BP.
XX AC
XX AAD61561;
XX DT
XX 15-JAN-2004 (first entry)
XX DE
XX Human GliTEN EST clone cDNA #2.
XX KW
XX Human; glioblastoma multiforme; GBM; GliTEN; brain cancer; diagnosis;
XX KW gene therapy; expressed sequence tag; ss.
XX OS
XX Homo sapiens.
XX PN
XX US2003108915-A1.
XX PD
XX 12-JUN-2003.
XX PF
XX 20-AUG-2002; 2002US-00224624.
XX PR
XX 20-OCT-2000; 2000US-0242160P.
XX PR
XX 20-OCT-2001; 2001US-00051769.
XX PA
XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX PI
XX McKinnon RD;
XX WPI; 2003-810848/76.
XX DR
XX Novel glioblastoma multiforme associated protein GliTEN useful for
XX PT treating glioblastoma multiforme and diagnosing brain cancer.
XX PS
XX Example 2; Page 8; Opp; English.
XX CC
XX The present invention provides novel glioblastoma multiforme (GBM)
XX CC associated protein GliTEN useful for treating glioblastoma multiforme and
XX CC diagnosing brain cancer. The invention is useful in brain cancer therapy,
XX CC treatment and diagnosis. The invention is also useful in gene therapy.
XX CC The present sequence is human GliTEN EST clone cDNA. This EST clone is
XX CC used in the exemplification of the invention.
XX SQ
XX Sequence 263 BP; 41 A; 78 C; 94 G; 50 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 22; DB 9; Length 263;
XX Best Local Similarity 100.0%; Pred. No. 1.1;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 GTGGAAGCCGCCGTGTACTCC 22
XX DB |||||||||||||||||||
XX 244 GTGGAAGCCGCCGTGTACTCC 223
XX
XX RESULT 5
XX ABX95679/c
XX ID ABX95679 standard; cDNA; 261 BP.
XX AC
XX ABX95679;
XX DT
XX 30-JUN-2003 (first entry)
XX DE
XX Human GliTEN expressed sequence tag.
XX KW
XX Human; ss; EST; chromosome 10q25; glioblastoma multiforme; brain cancer;
XX KW GliTEN; GBM; expressed sequence tag.
XX OS
XX Homo sapiens.
XX PN
XX US2003044811-A1.
XX PD
XX 06-MAR-2003.
XX PF
XX 20-OCT-2001; 2001US-00051769.
```

```
XX PR
XX 20-OCT-2000; 2000US-0242160P.
XX PA
XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX PI
XX McKinnon RD;
XX WPI; 2003-418995/39.
XX DR
XX New isolated nucleic acid representing a gene product associated with
XX PT Glioblastoma Multiforme, designated as GliTEN, useful as a molecular
XX PT marker for tumor cell identification and classification.
XX XX
XX Claim 1; Page 7; l1pp; English.
XX CC
XX The invention relates to an isolated nucleic acid representing a gene
XX CC product associated with Glioblastoma Multiforme, designated as GliTEN,
XX CC comprising the human EST (expressed sequence tag) appearing as ABX95678,
XX CC or a sequence that hybridises under stringent conditions to the EST, or
XX CC its complement. Also included are a probe for use in identifying a
XX CC patient at risk for progression into the malignant phenotype, comprising
XX CC the nucleic acids detailed above, detecting whether a patient is at risk
XX CC for progression into Glioblastoma Multiforme (comprising: (a) providing a
XX CC sample from a patient; (b) adding the probe to the sample or performing a
XX CC PCR analysis using the primers appearing as ABX95682 and ABX95683; (c)
XX CC analysing levels of mRNA bound with the probe; and (d) treating a control
XX CC sample to assess the level of mRNA in a control sample, where the
XX CC presence of increased levels of mRNA expression in the sample in an
XX CC amount higher than the control sample indicates risk for progression into
XX CC Glioblastoma Multiforme), and a kit for use in detecting whether a
XX CC patient is at risk for progression into Glioblastoma Multiforme
XX CC (comprising the probe or primers). The probe or the kit is useful for
XX CC detecting whether a patient is at risk for progression into Glioblastoma
XX CC Multiforme (a type of brain cancer). The EST useful as molecular marker
XX CC for tumour cell identification and classification and for diagnosing or
XX CC identifying candidates at risk for progression into a malignant phenotype
XX CC especially in brain cancer therapy, treatment and diagnosis. The present
XX CC sequence is the human GliTEN expressed sequence tag
XX SQ
XX Sequence 261 BP; 40 A; 81 C; 102 G; 38 T; 0 U; 0 Other;
XX
XX Query Match 92.7%; Score 20.4; DB 7; Length 261;
XX Best Local Similarity 95.5%; Pred. No. 6.4;
XX Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 GTGGAAGCCGCCGTGTACTCC 22
XX DB ||||||| ||||||| ||||||| |||||||
XX 244 GTGGAAGCCGCCGTGTACTCC 223
XX
XX RESULT 6
XX AAD61560/c
XX ID AAD61560 standard; cDNA; 261 BP.
XX AC
XX AAD61560;
XX DT
XX 15-JAN-2004 (first entry)
XX DE
XX Human GliTEN EST clone cDNA #1.
XX KW
XX Human; glioblastoma multiforme; GBM; GliTEN; brain cancer; diagnosis;
XX KW gene therapy; expressed sequence tag; ss.
XX OS
XX Homo sapiens.
XX PN
XX US2003108915-A1.
XX PD
XX 12-JUN-2003.
XX PF
XX 20-AUG-2002; 2002US-00224624.
XX PR
XX 20-OCT-2000; 2000US-0242160P.
XX PR
XX 20-OCT-2001; 2001US-00051769.
```

```
XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX
XX Mckinnon RD;
XX
XX WPI; 2003-810848/76.
XX
XX Novel glioblastoma multiforme associated protein GliTEN useful for
XX treating glioblastoma multiforme and diagnosing brain cancer.
XX
XX Claim 38; Page 11; Opp; English.
XX
XX The present invention provides novel glioblastoma multiforme (GBM)
XX associated protein GliTEN useful for treating glioblastoma multiforme and
XX diagnosing brain cancer. The invention is useful in brain cancer therapy,
XX treatment and diagnosis. The invention is also useful in gene therapy.
XX The present sequence is human GliTEN EST clone cDNA. This EST clone is
XX used in the exemplification of the invention
XX
XX Sequence 261 BP; 40 A; 81 C; 102 G; 38 T; 0 U; 0 Other;
XX
XX Query Match 92.7%; Score 20.4; DB 9; Length 261;
XX Best Local Similarity 95.5%; Pred. No. 6.4;
XX Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 GTGGAAGCGCGGTGTACTCC 22
XX ||||||| ||||||| |||||||
XX Db 244 GTGGAAGCGCGGTGTACTCC 223
XX
XX RESULT 7
XX ABX95680/c
XX ID ABX95680 standard; DNA; 873 BP.
XX
XX AC ABX95680;
XX
XX DT 30-JUN-2003 (first entry)
XX
XX DE Human GliTEN open reading frame.
XX
XX KW Human; ds; chromosome 10q25; glioblastoma multiforme; brain cancer;
XX GliTEN; GMB.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX CDS 1..873
XX /*tag= a
XX /partial
XX /product= "GliTEN"
XX /note= "No stop codon shown"
XX
XX PN US2003044811-A1.
XX
XX PD 06-MAR-2003.
XX
XX PF 20-OCT-2001; 2001US-00051769.
XX
XX PR 20-OCT-2000; 2000US-0242160P.
XX
XX PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX
XX PI Mckinnon RD;
XX
XX DR WPI; 2003-418995/39.
XX
XX New isolated nucleic acid representing a gene product associated with
XX Glioblastoma Multiforme, designated as GliTEN, useful as a molecular
XX marker for tumor cell identification and classification.
XX
XX Disclosure; Page 7; 11pp; English.
XX
XX The invention relates to an isolated nucleic acid representing a gene
XX
XX product associated with Glioblastoma Multiforme, designated as GliTEN,
XX comprising the human EST (expressed sequence tag) appearing as ABX95678,
XX or a sequence that hybridises under stringent conditions to the EST, or
XX its complement. Also included are a probe for use in identifying a
XX patient at risk for progression into the malignant phenotype, comprising
XX the nucleic acids detailed above, detecting whether a patient is at risk
XX for progression into Glioblastoma Multiforme (comprising: (a) providing a
XX sample from a patient; (b) adding the probe to the sample or performing a
XX PCR analysis using the primers appearing as ABX95682 and ABX95683; (c)
XX analysing levels of mRNA bound with the probe; and (d) treating a control
XX sample to assess the level of mRNA in a control sample, where the
XX presence of increased levels of mRNA expression in the sample in an
XX amount higher than the control sample indicates risk for progression into
XX Glioblastoma Multiforme), and a kit for use in detecting whether a
XX patient is at risk for progression into Glioblastoma Multiforme
XX (comprising the probe or primers). The probe or the kit is useful for
XX detecting whether a patient is at risk for progression into Glioblastoma
XX Multiforme (a type of brain cancer). The EST useful as molecular marker
XX for tumour cell identification and classification and for diagnosing or
XX identifying candidates at risk for progression into a malignant phenotype
XX especially in brain cancer therapy, treatment and diagnosis. The present
XX sequence is the open reading frame from the GliTEN gene
XX
XX Sequence 873 BP; 137 A; 304 C; 279 G; 153 T; 0 U; 0 Other;
XX
XX Query Match 92.7%; Score 20.4; DB 7; Length 873;
XX Best Local Similarity 95.5%; Pred. No. 6.8;
XX Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 GTGGAAGCGCGGTGTACTCC 22
XX ||||||| ||||||| |||||||
XX Db 609 GTGGAAGCGCGGTGTACTCC 588
XX
XX RESULT 8
XX ABX95678/c
XX ID ABX95678 standard; DNA; 960 BP.
XX
XX AC ABX95678;
XX
XX DT 30-JUN-2003 (first entry)
XX
XX DE Human gene encoding GliTEN.
XX
XX KW Human; ds; gene; chromosome 10q25; glioblastoma multiforme; brain cancer;
XX GliTEN; GMB.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX CDS 30..906
XX /*tag= a
XX /product= "GliTEN"
XX
XX PN US2003044811-A1.
XX
XX PD 06-MAR-2003.
XX
XX PF 20-OCT-2001; 2001US-00051769.
XX
XX PR 20-OCT-2000; 2000US-0242160P.
XX
XX PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX
XX PI Mckinnon RD;
XX
XX DR WPI; 2003-418995/39.
XX
XX New isolated nucleic acid representing a gene product associated with
XX Glioblastoma Multiforme, designated as GliTEN, useful as a molecular
XX marker for tumor cell identification and classification.
XX
XX Example 2; Fig 1; 11pp; English.
XX
```


XX The invention relates to an isolated nucleic acid representing a gene
 CC product associated with Glioblastoma Multiforme, designated as GLITEN,
 CC comprising the human EST (expressed sequence tag) appearing as ABX95678,
 CC or a sequence that hybridises under stringent conditions to the EST, or
 CC its complement. Also included are a probe for use in identifying a
 CC patient at risk for progression into the malignant phenotype, comprising
 CC the nucleic acids detailed above, detecting whether a patient is at risk
 CC for progression into Glioblastoma Multiforme (comprising: (a) providing a
 CC sample from a patient; (b) adding the probe to the sample or performing a
 CC PCR analysis using the primers appearing as ABX95682 and ABX95683; (c)
 CC analysing levels of mRNA bound with the probe; and (d) treating a control
 CC sample to assess the level of mRNA in a control sample, where the
 CC presence of increased levels of mRNA expression in the sample in an
 CC amount higher than the control sample indicates risk for progression into
 CC Glioblastoma Multiforme), and a kit for use in detecting whether a
 CC patient is at risk for progression into Glioblastoma Multiforme
 CC (comprising the probe or primers). The probe or the kit is useful for
 CC detecting whether a patient is at risk for progression into Glioblastoma
 CC Multiforme (a type of brain cancer). The EST useful as molecular marker
 CC for tumour cell identification and classification and for diagnosing or
 CC identifying candidates at risk for progression into a malignant phenotype
 CC especially in brain cancer therapy, treatment and diagnosis. The present
 CC sequence is a region of human chromosome 10 (q25) containing the GLITEN
 CC gene
 XX
 SQ Sequence 960 BP; 146 A; 324 C; 329 G; 161 T; 0 U; 0 Other;
 Query Match 92.7%; Score 20.4; DB 7; Length 960;
 Best Local Similarity 95.5%; Pred. No. 6.9;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GTGGAAGCCCGCTGTACTCC 22
 ||||| ||||| ||||| |||||
 Db 639 GTGGAAGCCCGCTGTACTCC 618
 RESULT 9
 AAD61559/c
 ID AAD61559 standard; cDNA; 960 BP.
 XX
 AC AAD61559;
 XX
 DT 15-JAN-2004 (first entry)
 XX
 DE Human GLITEN partial cDNA.
 XX
 KW Human; glioblastoma multiforme; GBM; GLITEN; brain cancer; diagnosis;
 KW gene therapy; chromosome 10; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN US2003108915-A1.
 XX
 PD 12-JUN-2003.
 XX
 PF 20-AUG-2002; 2002US-00224624.
 XX
 PR 20-OCT-2000; 2000US-0242160P.
 PR 20-OCT-2001; 2001US-00051769.
 XX
 PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
 XX
 PI McKinnon RD;
 XX
 DR WPI; 2003-810848/76.
 XX
 PT Novel glioblastoma multiforme associated protein GLITEN useful for
 PT treating glioblastoma multiforme and diagnosing brain cancer.
 XX
 PS Example 2; Fig 1; Opp; English.
 XX
 SQ The present invention provides novel glioblastoma multiforme (GBM)

CC associated protein GLITEN useful for treating glioblastoma multiforme and
 CC diagnosing brain cancer. The invention is useful in brain cancer therapy,
 CC treatment and diagnosis. The invention is also useful in gene therapy.
 CC The present sequence is human GLITEN partial cDNA. The human GLITEN gene
 CC is located on chromosome 10. This partial cDNA is used in the
 CC exemplification of the invention
 XX
 SQ Sequence 960 BP; 146 A; 324 C; 329 G; 161 T; 0 U; 0 Other;
 Query Match 92.7%; Score 20.4; DB 9; Length 960;
 Best Local Similarity 95.5%; Pred. No. 6.9;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GTGGAAGCCCGCTGTACTCC 22
 ||||| ||||| ||||| |||||
 Db 639 GTGGAAGCCCGCTGTACTCC 618
 RESULT 10
 ADE07789/c
 ID ADE07789 standard; DNA; 1092 BP.
 XX
 AC ADE07789;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Novel coding sequence (useful for identifying genetic disorders) #855.
 XX
 KW novel gene; novel protein; tissue marker; molecular weight marker;
 KW chromosome marker; genetic disorder; gene; ds.
 XX
 OS Unidentified.
 XX
 PN WO2003054152-A2.
 XX
 PD 03-JUL-2003.
 XX
 PF 10-DEC-2002; 2002WO-US039555.
 XX
 PR 10-DEC-2001; 2001US-0339739P.
 PR 11-DEC-2001; 2001US-0339453P.
 PR 14-MAR-2002; 2002US-0365091P.
 PR 14-MAR-2002; 2002US-0365384P.
 PR 12-APR-2002; 2002US-0372381P.
 PR 12-APR-2002; 2002US-0372615P.
 PR 22-APR-2002; 2002US-00128558.
 PR 24-APR-2002; 2002US-0376045P.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
 PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;
 PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;
 XX
 DR WPI; 2003-569235/53.
 DR P-PSDB; ADE08700.
 XX
 PT New polynucleotides, useful for expressing recombinant proteins for
 PT analysis, characterization or therapeutic use, or as markers for tissues
 PT in which the corresponding protein is preferentially expressed.
 XX
 PS Claim 1; SEQ ID NO 855; 1177pp; English.
 XX
 CC The invention comprises the amino acid and coding sequences of novel
 CC proteins. The DNA and protein sequences of the invention are useful as:
 CC markers for tissues in which the corresponding protein is preferentially
 CC expressed; as molecular weight markers on gels; as chromosome markers or
 CC tags; to identify chromosomes or to map related gene positions; and to
 CC compare with endogenous DNA sequences in patients to identify potential
 CC genetic disorders. The present DNA sequence represents a gene of the
 CC invention.
 XX
 SQ Sequence 1092 BP; 190 A; 377 C; 322 G; 203 T; 0 U; 0 Other;

Query Match 92.7%; Score 20.4; DB 9; Length 1092;
 Best Local Similarity 95.5%; Pred. No. 6.9;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGGAAGCGCGGTGTACTCC 22
 |||||
 DB 609 GTGGAAGCGCGGTGTACTCC 588

RESULT 11
 ADB62530/c
 ID ADB62530 standard; cDNA; 2639 BP.
 XX
 AC ADB62530;
 DT 04-DEC-2003 (first entry)
 XX Human cDNA encoding clone HHDPC20088160.
 DE
 XX Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
 KW tissue regeneration; cell regeneration; membrane protein;
 KW signal transduction-related protein; transcription-related protein;
 KW osteoporosis; neurological disease; cancer; tumour.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 214..2639
 FT /tag= a
 FT /partial
 FT /product= "Clone HHDPC20088160 protein"
 FT
 XX EF1308459-A2.
 XX
 XX 07-MAY-2003.
 XX
 XX 28-MAR-2002; 2002EP-00007401.
 XX
 XX 05-NOV-2001; 2001JP-00379298.
 PR 25-JAN-2002; 2002US-00350978.
 XX
 XX (HELI-) HELIX RES INST.
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Tamechika I;
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
 XX
 DR WPI: 2003-450961/43.
 DR P-PSDB; ADB64500.
 XX
 XX New polynucleotides and polypeptides, useful for developing a diagnostic
 marker or medicines for regulation of their expression and activity, or
 as targets of gene therapy.
 XX
 XX Claim 1; Page; 222pp; English.
 PS
 XX The invention discloses a polynucleotide comprising a sequence selected
 from 1970 fully defined nucleotide sequences which encode novel
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
 CC or its partial peptide, an antibody binding to the polypeptide or peptide
 CC of the polynucleotide, immunologically assaying the polypeptide or
 CC peptide of the polynucleotide by contacting the polypeptide or peptide
 CC with the antibody of the encoded protein, and observing the binding
 CC between the two, a transformant carrying the polynucleotide in an
 CC expressible manner and an antisense polynucleotide. The oligonucleotide
 CC is useful as a primer for synthesizing the polynucleotide, or as a probe
 CC for detecting the polynucleotide. The polynucleotides and encoded
 CC proteins are useful as pharmaceutical agents and many disease-related
 CC genes may be included in them, for developing a diagnostic marker or
 CC medicines for regulation of their expression and activity, or as targets
 CC of gene therapy. The genes are involved in tissue and/or cell

CC regeneration. Membrane proteins, signal transduction-related proteins,
 CC transcription-related proteins, disease-related proteins and genes
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
 CC neurological diseases, cancer, tumours). The cDNA may be used to regulate
 CC the activity or expression of the encoded protein to treat diseases. The
 CC sequence presented is a cDNA of the invention. Note: Some of the sequence
 CC data for this patent is not represented in the printed specification, but
 CC is based on sequence information supplied by the European Patent Office.
 XX
 SQ Sequence 2639 BP; 657 A; 702 C; 719 G; 561 T; 0 U; 0 Other;

Query Match 92.7%; Score 20.4; DB 9; Length 2639;
 Best Local Similarity 95.5%; Pred. No. 7.3;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGGAAGCGCGGTGTACTCC 22
 |||||
 DB 822 GTGGAAGCGCGGTGTACTCC 801

RESULT 12
 AAD61564/c
 ID AAD61564 standard; cDNA; 3832 BP.
 XX
 AC AAD61564;
 XX
 DT 15-JAN-2004 (first entry)
 XX
 XX Human GliTEN cDNA.
 DE
 XX Human; glioblastoma multiforme; GBM; GliTEN; brain cancer; diagnosis;
 KW gene therapy; chromosome 10; gene; ss.
 KW
 XX Homo sapiens.

XX Key Location/Qualifiers
 FH CDS 178..3642
 FT /tag= a
 FT /product= "Human GliTEN protein"
 FT /note= "The CDS is specifically claimed in claim 1"
 FT primer_bind 1011..11032
 FT /tag= b
 FT primer_bind 1405..1429
 FT /bound_moiety= "5' primer #843"
 FT /tag= c
 FT primer_bind 2083..2108
 FT /bound_moiety= "5' primer #1405"
 FT /tag= d
 FT primer_bind 2356..2378
 FT /bound_moiety= "5' primer #2083"
 FT /tag= e
 FT primer_bind 3625..3650
 FT /bound_moiety= "5' primer #2355"
 FT /tag= f
 FT primer_bind 3625..3650
 FT /bound_moiety= "3' primer #3636"

US2003108915-A1.
 XX
 PN 12-JUN-2003.
 XX
 PD 20-AUG-2002; 2002US-00224624.
 XX
 PF 20-OCT-2000; 2000US-0242160P.
 XX
 PR 20-OCT-2001; 2001US-00051769.
 XX
 XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
 PA
 XX McKinnon RD;
 PI
 XX WPI: 2003-810848/76.
 DR P-PSDB; AEW00758.
 XX
 XX Novel glioblastoma multiforme associated protein GliTEN useful for

PT treating glioblastoma multiforme and diagnosing brain cancer.

XX Claim 1; Fig 4; Opp; English.

XX The present invention provides novel glioblastoma multiforme (GBM)
CC associated protein GliTEN useful for treating glioblastoma multiforme and
CC diagnosing brain cancer. The invention is useful in brain cancer therapy,
CC treatment and diagnosis. The invention is also useful in gene therapy.
CC The present sequence is human GliTEN cDNA. The human GliTEN gene is
CC located on chromosome 10

XX
SQ Sequence 3832 BP; 1090 A; 926 C; 957 G; 859 T; 0 U; 0 Other;

Query Match 92.7%; Score 20.4; DB 9; Length 3832;
Best Local Similarity 95.5%; Pred.No. 7.4;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGGAAGCCCGCTGTGTACTCC 22
||||||| ||||||| ||||||| |||||||
Db 786 GTGGAAGCCCGCTGTGTACTCC 765

RESULT 13
ABK87764/c
ID ABK87764 standard; cDNA; 3847 BP.

XX ABK87764;

XX 24-SEP-2002 (first entry)

XX Human cDNA encoding glioma antigen KU-GB-5.

XX Human; ss; gene; glioma antigen; KU-GB-5; glioma; cytostatic; cancer;
KW immunostimulant; immunosuppressive; peptide therapy; gene therapy;
KW malignant brain tumour.

XX Homo sapiens.

XX Key Location/Qualifiers
FH 211..2883
FT CDS /*tag= a
FT /*product= "KU-GB-5"

XX WO200255695-A1.

XX 18-JUL-2002.

XX 30-NOV-2001; 2001WO-JP010505.

XX 09-JAN-2001; 2001JP-00001965.

XX (UYKE-) UNIV KEIO.

XX Toda M, Kawakami Y, Kawase T, Iizuka Y;

XX WPI; 2002-538141/57.

XX P-PSDB; AAU99614.

XX New human glioma antigen for diagnosing and treating glioma, human
PT malignant brain tumor and other cancers, and for studying onset of
PT glioma.

XX Claim 12; Page 94-100; 109pp; Japanese.

XX The invention relates to preparing glioma antigen and/or glioma antigen
CC gene comprises e.g. extracting and isolating total RNA from a glioma cell
CC line then synthesising cDNA for constructing a lambda phage cDNA library
CC and reacting the library with the serum for reaction and detecting
CC positive clones reacting with the antibody in the serum by using a
CC labeled anti-immunoglobulin (Ig)G antibody. Also included are diagnostic
CC reagents for detecting glioma containing 1 or more kinds of the whole or
CC partial glioma antigens thus prepared, and/or an antibody binding with
CC the glioma antigens and/or parts of them, probes for detecting or

CC diagnosis of glioma containing the whole or partial antisense chains of
CC DNAs or RNAs encoding the prepared whole or partial glioma antigens, DNAs
CC (or cDNAs) encoding the glioma antigens (AAU99609-AAU99614) and the
CC protein sequences for the antigens (or peptides derived from them), a DNA
CC hybridisable with the DNA in under stringent conditions and encoding a
CC protein with immune induction activity, a fusion protein or peptide
CC formed by bonding the protein or its partial peptide with a marker
CC protein or peptide tag, an antibody for the protein or its partial
CC peptide, a host cell containing an expression system for expressing the
CC protein or peptide, an non-human animal which has deletion of the gene
CC function encoding the protein or its partial peptide on the chromosome,
CC or has overexpression of the protein or its partial peptide and screening
CC substances promoting or inhibiting immune induction activity by using the
CC protein or its partial peptide, the test substances and T cells, and
CC measuring and evaluating immune induction activity in T cells. The
CC antigen and DNA encoding it, are applicable in diagnosis and treatment of
CC (e.g. by peptide or gene therapy) glioma, human malignant brain tumour
CC and other cancers, and for studying onset of glioma. The present sequence
CC encodes the glioma antigen KU-GB-5

XX Sequence 3847 BP; 1197 A; 744 C; 819 G; 1086 T; 0 U; 1 Other;

Query Match 92.7%; Score 20.4; DB 6; Length 3847;
Best Local Similarity 95.5%; Pred.No. 7.4;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGGAAGCCCGCTGTGTACTCC 22
||||||| ||||||| ||||||| |||||||
Db 30 GTGGAAGCCCGCTGTGTACTCC 9

RESULT 14
AAKS1608/c
ID AAKS1608 standard; cDNA; 4450 BP.

XX AAKS1608;

XX 06-NOV-2001 (first entry)

XX Human polynucleotide SEQ ID NO 153.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.

XX Homo sapiens.

XX WO200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US004098.

XX 03-FEB-2000; 2000US-00496914.

XX 27-APR-2000; 2000US-00560875.

XX 20-JUN-2000; 2000US-00598075.

XX 19-JUL-2000; 2000US-00620325.

XX 01-SEP-2000; 2000US-00654936.

XX 15-SEP-2000; 2000US-00663561.

XX 20-OCT-2000; 2000US-00693325.

XX 30-NOV-2000; 2000US-00728422.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;

XX Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WPI; 2001-476283/51.

XX P-PSDB; AAM78475.

XX Nucleic acids encoding polypeptides with cytokine-like activities, useful

PT in diagnosis and gene therapy.
XX
PS Claim 1; Page 862-867; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 4450 BP; 1252 A; 987 C; 1047 G; 1164 T; 0 U; 0 Other;
Query Match 92.7%; Score 20.4; DB 4; Length 4450;
Best Local Similarity 95.5%; Pred. No. 7.5;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GTGGAAGCGCGGTGTACTCC 22
Db 609 GTGGAGGCCCGGTGTACTCC 588
RESULT 15
ABA09174/c
ID ABA09174 standard; cDNA; 4470 BP.
XX ABA09174;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human secreted protein homologue-encoding cDNA, SEQ ID NO:950.
XX
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytotatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnary; antitumor; ss.
XX
OS Homo sapiens.
XX
PN WO200157188-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US003800.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-457740/49.
DR P-PSDB; ABB11930.
XX
PT Human proteins and DNA encoding sequences useful for preventing, treating
PT or ameliorating a medical condition in a mammalian subject e.g. arthritis

PT and cancer.
XX
PS Claim 1; Page 813-814; 1963pp; English.
XX
CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, and hence
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haematopoiesis regulatory activity; activin- or inhibin-related activities;
CC immunomodulatory activity; activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g., by protein or gene therapy. Such conditions include
CC cancers, haematopoietic disorders (e.g. myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a cDNA encoding a
CC novel human polypeptide of the invention
XX
SQ Sequence 4470 BP; 1254 A; 991 C; 1054 G; 1171 T; 0 U; 0 Other;
Query Match 92.7%; Score 20.4; DB 4; Length 4470;
Best Local Similarity 95.5%; Pred. No. 7.5;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GTGGAAGCGCGGTGTACTCC 22
Db 609 GTGGAGGCCCGGTGTACTCC 588
Search completed: July 18, 2004, 12:45:05
Job time : 37.7148 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 12:32:20 ; Search time 7.42951 Seconds
(without alignments)
1643.301 Million cell updates/sec

Title: US-10-051-769-6
Perfect score: 22
Sequence: 1 gtggaagccgcgttgactcc 22

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgm2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgm2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgm2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgm2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgm2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgm2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	20.4	92.7	4303	4	US-09-976-594-899
c 2	18.8	85.5	960	2	US-07-841-591A-1
c 3	18.8	85.5	960	5	PCT-US93-02034-1
c 4	17.2	78.2	681	2	US-08-347-594A-3
c 5	17.2	78.2	681	3	US-08-463-682-18
c 6	17.2	78.2	750	2	US-08-797-689-3
c 7	17.2	78.2	750	4	US-09-984-186-3
c 8	17.2	78.2	1818	4	US-09-252-991A-2266
c 9	17.2	78.2	3051	4	US-09-252-991A-2343
c 10	17.2	78.2	3156	4	US-09-252-991A-2574
c 11	17.2	78.2	3234	1	US-08-264-534-31
c 12	17.2	78.2	3234	1	US-08-083-590A-10
c 13	17.2	78.2	3234	1	US-08-465-500-31
c 14	17.2	78.2	3234	2	US-08-346-128-31
c 15	17.2	78.2	3234	3	US-08-532-384-10
c 16	17.2	78.2	3234	3	US-08-893-828-31
c 17	17.2	78.2	6153	2	US-08-347-594A-1
c 18	17.2	78.2	6153	3	US-08-463-682-2
c 19	17.2	78.2	8575	4	US-09-381-261A-2
c 20	17.2	78.2	8802	3	US-08-896-449A-1
c 21	17.2	78.2	8802	3	US-09-132-652-1
c 22	16.4	74.5	2064	1	US-08-343-428-1
c 23	16.2	73.6	237	4	US-09-328-352-888
c 24	16.2	73.6	441	4	US-09-252-991A-5335
c 25	16.2	73.6	486	4	US-09-252-991A-13723
c 26	16.2	73.6	693	4	US-09-252-991A-3399
c 27	16.2	73.6	798	4	US-09-252-991A-3359

c 28	16.2	73.6	843	4	US-09-328-352-849	Sequence 849, Appl
c 29	16.2	73.6	1041	4	US-09-252-991A-5300	Sequence 5300, Ap
c 30	16.2	73.6	1341	4	US-09-252-991A-5263	Sequence 5263, Ap
c 31	16.2	73.6	1452	4	US-09-252-991A-3313	Sequence 3313, Ap
c 32	16.2	73.6	1867	3	US-08-943-731-111	Sequence 111, Appl
c 33	16.2	73.6	1870	2	US-08-359-705B-3	Sequence 3, Appli
c 34	16.2	73.6	1870	2	US-08-286-846A-3	Sequence 3, Appli
c 35	16.2	73.6	1870	2	US-08-457-880A-3	Sequence 3, Appli
c 36	16.2	73.6	1870	3	US-08-444-622A-3	Sequence 3, Appli
c 37	16.2	73.6	1870	3	US-08-942-562-3	Sequence 3, Appli
c 38	16.2	73.6	1870	3	US-09-156-923-3	Sequence 13526, A
c 39	16.2	73.6	2496	4	US-09-252-991A-13526	Sequence 13526, A
c 40	16.2	73.6	2622	4	US-09-252-991A-13933	Sequence 2993, Ap
c 41	16.2	73.6	2673	4	US-09-252-991A-2993	Sequence 13933, A
c 42	16.2	73.6	2685	4	US-09-252-991A-13826	Sequence 2686, Ap
c 43	16.2	73.6	2715	4	US-09-252-991A-2686	Sequence 2686, Ap
c 44	16.2	73.6	2814	4	US-09-252-991A-2875	Sequence 2875, Ap
c 45	16.2	73.6	3060	1	US-08-286-305A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-976-594-899/c
; Sequence 899, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 899
; LENGTH: 4303
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673549 225420.1
US-09-976-594-899

Query Match 92.7%; Score 20.4; DB 4; Length 4303;
Best Local Similarity 95.5%; Pred. No. 1.6;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGGAAGCCGCCGTGTACTCC 22
|||||
Db 322 GTGGAAGCCGCCGTGTACTCC 301

RESULT 2
US-07-841-591A-1/c
; Sequence 1, Application US/07841591A
; Patent No. 5900476
; GENERAL INFORMATION:
; APPLICANT: Ruggeri, Zaverio M. and
; APPLICANT: Ware, Jerry, inventors
; APPLICANT: on behalf of The Scripps Research
; APPLICANT: Institute
; TITLE OF INVENTION: Therapeutic Domains of
; TITLE OF INVENTION: von Willebrand Factor
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute
; STREET: 10666 No. 5900476th Torrey Pines Road
; CITY: La Jolla
; STATE: California

;; COUNTRY: United States
;; ZIP: 92037
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: 1.2 megabyte 5 1/4" floppy
;; COMPUTER: AST Bravo 386SX
;; OPERATING SYSTEM: MS DOS version 3.2
;; SOFTWARE: WordPerfect 5.1 conv. to ASCII
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/841,591A
;; FILING DATE: 26-Feb-92
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA: This is a c-i-p of
;; APPLICATION NUMBER: PCT/US91/07756
;; FILING DATE: 17-Oct-91
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Barron, Alexis
;; REGISTRATION NUMBER: 22,702
;; REFERENCE/DOCKET NUMBER: P16,633-H
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (215) 923-4466
;; TELEFAX: (215) 923-2189
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 960
;; TYPE: Nucleic Acid
;; STRANDEDNESS: single stranded
;; TOPOLOGY: linear
;; US-07-841-591A-1

Query Match 85.5%; Score 18.8; DB 2; Length 960;
Best Local Similarity 90.9%; Pred. No. 7.8;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTGGAAGCCGCGTGTACTCC 22
DB 399 GTGGAAGCCGCGTGTACTCC 378

RESULT 3
PCT-US93-02034-1/c
;; Sequence 1, Application PC/TUS9302034
;; GENERAL INFORMATION:
;; APPLICANT: Ruggeri, Zaverio M. and
;; APPLICANT: Ware, Jerry, inventors
;; APPLICANT: on behalf of The Scripps Research
;; APPLICANT: Institute
;; TITLE OF INVENTION: Therapeutic Domains of
;; TITLE OF INVENTION: von Willebrand Factor
;; NUMBER OF SEQUENCES: 15
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: The Scripps Research Institute
;; STREET: 10666 North Torrey Pines Road
;; CITY: La Jolla
;; STATE: California
;; COUNTRY: United States
;; ZIP: 92037
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: 2 megabyte 3.25" floppy
;; COMPUTER: AST Bravo 386SX
;; OPERATING SYSTEM: MS DOS version 3.2
;; SOFTWARE: WordPerfect 5.1 conv. to ASCII
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US93/02034
;; FILING DATE: 19930223
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA: This is a c-i-p of
;; APPLICATION NUMBER: PCT/US91/07756
;; FILING DATE: 17-Oct-91
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Barron, Alexis
;; REGISTRATION NUMBER: 22,702
;; REFERENCE/DOCKET NUMBER: P16,633-H PCT
;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (215) 923-4466
;; TELEFAX: (215) 923-2189
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 960
;; TYPE: NUCLEIC ACID
;; STRANDEDNESS: single stranded
;; TOPOLOGY: linear
;; PCT-US93-02034-1

Query Match 85.5%; Score 18.8; DB 5; Length 960;
Best Local Similarity 90.9%; Pred. No. 7.8;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTGGAAGCCGCGTGTACTCC 22
DB 399 GTGGAAGCCGCGTGTACTCC 378

RESULT 4
US-08-347-594A-3/c
;; Sequence 3, Application US/08347594A
;; Patent No. 5849536
;; GENERAL INFORMATION:
;; APPLICANT: Garfinkel, Leonard
;; APPLICANT: Richter, Tamar
;; TITLE OF INVENTION: CLONING AND PRODUCTION OF HUMAN VON
;; TITLE OF INVENTION: WILLEBRAND FACTOR GPIIb BINDING DOMAIN POLYPEPTIDES AND
;; TITLE OF INVENTION: METHODS OF USING SAME
;; NUMBER OF SEQUENCES: 4
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: John P. White
;; STREET: 1185 Avenue of the Americas
;; CITY: New York
;; STATE: New York
;; COUNTRY: USA
;; ZIP: 10036
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/347,594A
;; FILING DATE: No. 5849536ember 30, 1994
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: White, John P.
;; REGISTRATION NUMBER: 28,678
;; REFERENCE/DOCKET NUMBER: 36537-B2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 212-278-0400
;; TELEFAX: 212-391-0525
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 681 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..681
;; US-08-347-594A-3

Query Match 78.2%; Score 17.2; DB 2; Length 681;
Best Local Similarity 86.4%; Pred. No. 42;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTGGAAGCCGCGTGTACTCC 22
DB 183 GTGGAAGCCGCGTGTACTCC 162

```
RESULT 5
US-08-463-682-18/c
; Sequence 18, Application US/08463682
; Patent No. 6008193
; GENERAL INFORMATION:
; APPLICANT: Leonard Garfinkel, et al.
; TITLE OF INVENTION: Cloning and Production of Human Von
; TITLE OF INVENTION: Willebrand Factor GPIb Binding Domain Polypeptides and
; TITLE OF INVENTION: Methods of Using Same
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,682
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 36537-B2-Y
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 681 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..681
US-08-463-682-18
Query Match 78.2%; Score 17.2; DB 3; Length 681;
Best Local Similarity 86.4%; Pred. No. 42;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GTGGAAGCCGCGTGTTACTCC 22
Db 183 GTGGAGCCGCGTGTTACTCC 162

RESULT 6
US-08-797-689-3/c
; Sequence 3, Application US/08797689
; Patent No. 5876969
; GENERAL INFORMATION:
; APPLICANT: Fleer, Reinhard
; APPLICANT: Fournier, Alain
; APPLICANT: Guitton, Jean-Dominique
; APPLICANT: Jung, Gerard
; APPLICANT: Yeh, Patrice
; TITLE OF INVENTION: NOVEL BIOLOGICALLY ACTIVE POLYPEPTIDES,
; TITLE OF INVENTION: PREPARATION THEREOF AND PHARMACEUTICAL COMPOSITION
; TITLE OF INVENTION: CONTAINING SAID POLYPEPTIDES
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Road, 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19426
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
STREET: 500 Arcola Road, 3C43
CITY: Collegeville
STATE: PA
COUNTRY: USA
ZIP: 19426
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: System 7.1
SOFTWARE: Word 5.1 (PatentIn)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/797,689
FILING DATE: 31-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/256,927
FILING DATE: 28-JUL-1994
APPLICATION NUMBER: FR 92/01064
FILING DATE: 31-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00085
FILING DATE: 28-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Smith Ph.D., Julie K.
REGISTRATION NUMBER: P-38,619
REFERENCE/DOCKET NUMBER: ST92006-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (610) 454-3839
TELEFAX: (610) 454-3808
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 750 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 3..746
US-08-797-689-3
Query Match 78.2%; Score 17.2; DB 2; Length 750;
Best Local Similarity 86.4%; Pred. No. 42;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GTGGAAGCCGCGTGTTACTCC 22
Db 293 GTGGAGCCGCGTGTTACTCC 272

RESULT 7
US-09-984-186-3/c
; Sequence 3, Application US/09984186
; Patent No. 6686179
; GENERAL INFORMATION:
; APPLICANT: Fleer, Reinhard
; APPLICANT: Fournier, Alain
; APPLICANT: Guitton, Jean-Dominique
; APPLICANT: Jung, Gerard
; APPLICANT: Yeh, Patrice
; TITLE OF INVENTION: NOVEL BIOLOGICALLY ACTIVE POLYPEPTIDES,
; TITLE OF INVENTION: PREPARATION THEREOF AND PHARMACEUTICAL COMPOSITION
; TITLE OF INVENTION: CONTAINING SAID POLYPEPTIDES
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Road, 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19426
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

COMPUTER: Macintosh
OPERATING SYSTEM: System 7.1
SOFTWARE: Word 5.1 (PatentIn)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/984,186
FILING DATE: 29-Oct-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/797,689
FILING DATE: 31-JAN-1997
APPLICATION NUMBER: US 08/256,927
FILING DATE: 28-JUL-1994
APPLICATION NUMBER: FR 92/01064
FILING DATE: 31-JAN-1992
APPLICATION NUMBER: PCT/FR93/00085
FILING DATE: 28-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Smith Ph.D., Julie K.
REGISTRATION NUMBER: P-38,619
REFERENCE/DOCKET NUMBER: ST92006-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (610) 454-3839
TELEFAX: (610) 454-3808
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 750 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 3..746
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-984-186-3
Query Match 78.2%; Score 17.2; DB 4; Length 750;
Best Local Similarity 86.4%; Pred. No. 42;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GTGGAAGCGCGGTGTACTCC 22
Db 293 GTGGAGCGCGGTGTACTCC 272
RESULT 8
US-09-252-991A-2266
Sequence 2266, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 2266
LENGTH: 1818
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2266
Query Match 78.2%; Score 17.2; DB 4; Length 1818;
Best Local Similarity 86.4%; Pred. No. 43;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GTGGAAGCGCGGTGTACTCC 22

Db 1024 GTGGAAGCGCGGTGTACTTC 1045
RESULT 9
US-09-252-991A-2343
Sequence 2343, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 2343
LENGTH: 3051
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
FEATURE:
NAME/KEY: unsure
LOCATION: (2880)
OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-252-991A-2343
Query Match 78.2%; Score 17.2; DB 4; Length 3051;
Best Local Similarity 86.4%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GTGGAAGCGCGGTGTACTCC 22
Db 1935 GTGGAAGCGCGGTGTACTTC 1956
RESULT 10
US-09-252-991A-2574/c
Sequence 2574, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 2574
LENGTH: 3156
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
FEATURE:
NAME/KEY: unsure
LOCATION: (1027)
OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-252-991A-2574
Query Match 78.2%; Score 17.2; DB 4; Length 3156;
Best Local Similarity 86.4%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GTGGAAGCGCGGTGTACTCC 22
Db 1972 GTGGAAGCGCGGTGTACTTC 1951


```
RESULT 11
US-08-264-534-31
; Sequence 31, Application US/08264534
; Patent No. 5648464
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon et al.
; TITLE OF INVENTION: Human No. 5648464ch And Delta, Binding Domains
; TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/264,534
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/695,189
; FILING DATE: 03-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3234
US-08-264-534-31

Query Match 78.2%; Score 17.2; DB 1; Length 3234;
Best Local Similarity 86.4%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTGGAAGCGCGCTGTGACTCC 22
|||||
Db 1762 GTGATGCCGCGAGTTGTCTCC 1783

RESULT 12
US-08-083-590A-10
; Sequence 10, Application US/08083590A
; Patent No. 5786158
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
; TITLE OF INVENTION: Nucleic Acids
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
```

```
ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3234
US-08-083-590A-10

Query Match 78.2%; Score 17.2; DB 1; Length 3234;
Best Local Similarity 86.4%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTGGAAGCGCGCTGTGACTCC 22
|||||
Db 1762 GTGATGCCGCGAGTTGTCTCC 1783

RESULT 13
US-08-465-500-31
; Sequence 31, Application US/08465500
; Patent No. 5789195
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Muskavitch, Marc A.T.
; APPLICANT: Fehon, Richard G.
; APPLICANT: Rebay, Ilaria
; APPLICANT: Blaumüller, Cristine M.
; APPLICANT: Shepard, Scott B.
; TITLE OF INVENTION: HUMAN NOTCH AND DELTA, BINDING DOMAINS
; TITLE OF INVENTION: IN TOPORYTHMIC PROTEINS, AND METHODS BASED THEREON
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,500
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-034
```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 3234 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-465-500-31

Query Match 78.2%; Score 17.2; DB 1; Length 3234;
Best Local Similarity 86.4%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTGGAAGCCGCGTGTACTCC 22
||||| ||||| ||||| ||||| |||||
Db 1762 GTGGATGCCGCGAGTTGTGCTCC 1783

RESULT 14
US-08-346-128-31
Sequence 31, Application US/08346128
Patent No. 5856441
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon et al.
TITLE OF INVENTION: Human No. 5856441ch And Delta, Binding Domains
TITLE OF INVENTION: In Toporychmic Proteins, And Methods Based Thereon
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/346,128
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/879,038
FILING DATE: 30-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 3234 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3234
US-08-346-128-31

Query Match 78.2%; Score 17.2; DB 2; Length 3234;
Best Local Similarity 86.4%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTGGAAGCCGCGTGTACTCC 22
||||| ||||| ||||| ||||| |||||
Db 1762 GTGGATGCCGCGAGTTGTGCTCC 1783

RESULT 15

US-08-532-384-10
Sequence 10, Application US/08532384
Patent No. 6083904
GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, S. et al.
TITLE OF INVENTION: Therapeutic And Diagnostic Methods
TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And
TITLE OF INVENTION: Nucleic Acids
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/532,384
FILING DATE:
CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/083,590
FILING DATE: 25-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-015
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 3234 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3234
US-08-532-384-10

Query Match 78.2%; Score 17.2; DB 3; Length 3234;
Best Local Similarity 86.4%; Pred. No. 44;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTGGAAGCCGCGTGTACTCC 22
||||| ||||| ||||| ||||| |||||
Db 1762 GTGGATGCCGCGAGTTGTGCTCC 1783

Search completed: July 18, 2004, 14:38:30
Job time : 8.42951 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 12:36:30 ; Search time 38.5902 Seconds
(without alignments)
2781.271 Million cell updates/sec

Title: US-10-051-769-6

Perfect score: 22
Sequence: 1 GTGGAAGCCGCGTGTACTCC 22

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3190392 seqs, 2439311697 residues

Total number of hits satisfying chosen parameters: 6381984

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PTCUT_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	22	15	US-10-051-769-6
2	22	100.0	22	15	US-10-224-624-6
3	22	100.0	263	15	US-10-051-769-4
4	22	100.0	263	15	US-10-224-624-4
5	20.4	92.7	261	15	US-10-051-769-2
6	20.4	92.7	261	15	US-10-224-624-2
7	20.4	92.7	873	15	US-10-051-769-3
8	20.4	92.7	873	15	US-10-224-624-3
9	20.4	92.7	960	15	US-10-051-769-1
10	20.4	92.7	960	15	US-10-224-624-1
11	20.4	92.7	2639	16	US-10-104-047-684
12	20.4	92.7	3465	15	US-10-224-624-9
13	20.4	92.7	3649	13	US-10-112-944-63
14	20.4	92.7	3832	15	US-10-224-624-7

c 15	20.4	92.7	3847	12	US-10-250-823-11	Sequence 11, Appl
c 16	20.4	92.7	4470	13	US-10-276-774-950	Sequence 950, Appl
c 17	20.4	92.7	4801	12	US-10-336-603A-25	Sequence 25, Appl
c 18	17.2	78.2	750	9	US-09-984-186-3	Sequence 3, Appl
c 19	17.2	78.2	750	15	US-10-237-667-3	Sequence 3, Appl
c 20	17.2	78.2	750	15	US-10-237-708-3	Sequence 3, Appl
c 21	17.2	78.2	750	15	US-10-237-866-3	Sequence 3, Appl
c 22	17.2	78.2	750	15	US-10-237-871-3	Sequence 3, Appl
c 23	17.2	78.2	750	15	US-10-237-624-3	Sequence 3, Appl
c 24	17.2	78.2	750	17	US-10-702-536-3	Sequence 3, Appl
c 25	17.2	78.2	750	17	US-10-702-636-3	Sequence 3, Appl
c 26	17.2	78.2	840	13	US-10-027-632-135878	Sequence 135878,
c 27	17.2	78.2	840	16	US-10-027-632-135878	Sequence 135878,
c 28	17.2	78.2	855	15	US-10-156-761-5908	Sequence 5908, Ap
c 29	17.2	78.2	1075	13	US-10-425-114-785	Sequence 785, App
c 30	17.2	78.2	1080	15	US-10-156-761-7141	Sequence 7141, Ap
c 31	17.2	78.2	1513	15	US-10-029-386-22889	Sequence 22889, A
c 32	17.2	78.2	1703	13	US-10-425-114-23185	Sequence 23185, A
c 33	17.2	78.2	1709	13	US-10-425-114-24394	Sequence 24394, A
c 34	17.2	78.2	2460	15	US-10-004-113-60	Sequence 60, Appl
c 35	17.2	78.2	3494	13	US-10-160-497-13	Sequence 13, Appl
c 36	17.2	78.2	3494	13	US-10-348-750-13	Sequence 13, Appl
c 37	17.2	78.2	6360	9	US-09-954-456-2256	Sequence 2256, Ap
c 38	17.2	78.2	7332	9	US-09-944-849-1	Sequence 1, Appl
c 39	17.2	78.2	7615	15	US-10-004-113-59	Sequence 59, Appl
c 40	17.2	78.2	7673	9	US-09-815-525-1	Sequence 1, Appl
c 41	17.2	78.2	7673	16	US-10-294-006-1	Sequence 1, Appl
c 42	17.2	78.2	7693	13	US-10-072-012-133	Sequence 133, App
c 43	17.2	78.2	7693	13	US-10-160-497-11	Sequence 11, Appl
c 44	17.2	78.2	7693	13	US-10-348-750-11	Sequence 11, Appl
c 45	17.2	78.2	8575	9	US-09-381-261A-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-10-051-769-6
; Sequence 6, Application US/10051769
; Publication No. US20030044811A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randy D.
; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION
; FILE REFERENCE: 268/260 (RMJ-00-37)
; CURRENT APPLICATION NUMBER: US/10/051,769
; PRIOR FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,160
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 22
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Primer
; US-10-051-769-6

Query Match 100.0%; Score 22; DB 15; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.28; Mismatches 0; Indels 0; Gaps 0;
Matches 22; Conservative 0;

Qy 1 GTGGAAGCCGCGTGTACTCC 22
|||||
Db 1 GTGGAAGCCGCGTGTACTCC 22

RESULT 2
US-10-224-624-6
; Sequence 6, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randall D.

```
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein Gliten
; FILE REFERENCE: 54704.8059.US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 6
; LENGTH: 22
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Primer
US-10-224-624-6

Query Match          100.0%; Score 22; DB 15; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.28; 0; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;

QY 1 GTGGAAGCCGCGTGTACTCC 22
Db 1 GTGGAAGCCGCGTGTACTCC 22

RESULT 3
US-10-051-769-4/c
; Sequence 4, Application US/10051769
; Publication No. US20030044811A1
; GENERAL INFORMATION:
; APPLICANT: McKINNON, Randy D.
; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION
; FILE REFERENCE: 268/260 (RWJ-00-37)
; CURRENT APPLICATION NUMBER: US/10/051,769
; CURRENT FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,160
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 4
; LENGTH: 263
; TYPE: DNA
; ORGANISM: Sprague Dawley rat
US-10-051-769-4

Query Match          100.0%; Score 22; DB 15; Length 263;
Best Local Similarity 100.0%; Pred. No. 0.28; 0; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;

QY 1 GTGGAAGCCGCGTGTACTCC 22
Db 1 GTGGAAGCCGCGTGTACTCC 22

RESULT 4
US-10-224-624-4/c
; Sequence 4, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: McKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein Gliten
; FILE REFERENCE: 54704.8059.US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 4
; LENGTH: 263
; TYPE: DNA
; ORGANISM: Sprague Dawley rat
US-10-051-769-4

Query Match          100.0%; Score 22; DB 15; Length 263;
Best Local Similarity 100.0%; Pred. No. 0.28; 0; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;

QY 1 GTGGAAGCCGCGTGTACTCC 22
Db 244 GTGGAAGCCGCGTGTACTCC 223

RESULT 5
US-10-051-769-2/c
; Sequence 2, Application US/10051769
; Publication No. US20030044811A1
; GENERAL INFORMATION:
; APPLICANT: McKINNON, Randy D.
; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION
; FILE REFERENCE: 268/260 (RWJ-00-37)
; CURRENT APPLICATION NUMBER: US/10/051,769
; CURRENT FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,160
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 261
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-051-769-2

Query Match          92.7%; Score 20.4; DB 15; Length 261;
Best Local Similarity 95.5%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGGAAGCCGCGTGTACTCC 22
Db 244 GTGGAAGCCGCGTGTACTCC 223

RESULT 6
US-10-224-624-2/c
; Sequence 2, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: McKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein Gliten
; FILE REFERENCE: 54704.8059.US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 261
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-224-624-2

Query Match          92.7%; Score 20.4; DB 15; Length 261;
Best Local Similarity 95.5%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGGAAGCCGCGTGTACTCC 22
Db 244 GTGGAAGCCGCGTGTACTCC 223
```

```
; LENGTH: 263
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-224-624-4

Query Match          100.0%; Score 22; DB 15; Length 263;
Best Local Similarity 100.0%; Pred. No. 0.28; 0; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;

QY 1 GTGGAAGCCGCGTGTACTCC 22
Db 244 GTGGAAGCCGCGTGTACTCC 223

RESULT 5
US-10-051-769-2/c
; Sequence 2, Application US/10051769
; Publication No. US20030044811A1
; GENERAL INFORMATION:
; APPLICANT: McKINNON, Randy D.
; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION
; FILE REFERENCE: 268/260 (RWJ-00-37)
; CURRENT APPLICATION NUMBER: US/10/051,769
; CURRENT FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,160
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 261
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-051-769-2

Query Match          92.7%; Score 20.4; DB 15; Length 261;
Best Local Similarity 95.5%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGGAAGCCGCGTGTACTCC 22
Db 244 GTGGAAGCCGCGTGTACTCC 223

RESULT 6
US-10-224-624-2/c
; Sequence 2, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: McKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein Gliten
; FILE REFERENCE: 54704.8059.US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 261
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-224-624-2

Query Match          92.7%; Score 20.4; DB 15; Length 261;
Best Local Similarity 95.5%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGGAAGCCGCGTGTACTCC 22
Db 244 GTGGAAGCCGCGTGTACTCC 223
```

```

; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-051-769-1

Query Match          92.7%; Score 20.4; DB 15; Length 960;
Best Local Similarity 95.5%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGGAAGCCGCGTGTACTCC 22
Db 639 GTGGAAGCCCCCGTGTACTCC 618

RESULT 10
US-10-224-624-1/c
; Sequence 1, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: McKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GliTEN
; FILE REFERENCE: 54704.8059.US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-224-624-1

Query Match          92.7%; Score 20.4; DB 15; Length 960;
Best Local Similarity 95.5%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGGAAGCCGCGTGTACTCC 22
Db 639 GTGGAAGCCCCCGTGTACTCC 618

RESULT 11
US-10-104-047-684/c
; Sequence 684, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 684
; LENGTH: 2639
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-684

Query Match          92.7%; Score 20.4; DB 16; Length 2639;
Best Local Similarity 95.5%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGGAAGCCGCGTGTACTCC 22
```

```

; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-051-769-1

Query Match          92.7%; Score 20.4; DB 15; Length 873;
Best Local Similarity 95.5%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGGAAGCCGCGTGTACTCC 22
Db 609 GTGGAAGCCCCCGTGTACTCC 588

RESULT 8
US-10-224-624-3/c
; Sequence 3, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: McKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GliTEN
; FILE REFERENCE: 54704.8059.US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 873
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-224-624-3

Query Match          92.7%; Score 20.4; DB 15; Length 873;
Best Local Similarity 95.5%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGGAAGCCGCGTGTACTCC 22
Db 609 GTGGAAGCCCCCGTGTACTCC 588

RESULT 9
US-10-051-769-1/c
; Sequence 1, Application US/10051769
; Publication No. US20030044811A1
; GENERAL INFORMATION:
; APPLICANT: McKINNON, Randy D.
; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION
; FILE REFERENCE: 268/260 (RWJ-00-37)
; CURRENT APPLICATION NUMBER: US/10/051,769
; CURRENT FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US/10/051,769
; PRIOR FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,160
; PRIOR FILING DATE: 2000-10-20
```

```
Db      822 GTGGAAGCCCCCGTTGTACTCC 801
|||||
RESULT 12
US-10-224-624-9/c
; Sequence 9, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GlitEN
; FILE REFERENCE: 54704.8059.US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 3465
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-224-624-9

Query Match          92.7%; Score 20.4; DB 15; Length 3465;
Best Local Similarity 95.5%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GTGGAAGCCCCCGTTGTACTCC 22
|||||

Db      609 GTGGAAGCCCCCGTTGTACTCC 588

RESULT 14
US-10-224-624-7/c
; Sequence 7, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GlitEN
; FILE REFERENCE: 54704.8059.US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 3832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (178)..(3639)
; OTHER INFORMATION:
US-10-224-624-7

Query Match          92.7%; Score 20.4; DB 15; Length 3832;
Best Local Similarity 95.5%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GTGGAAGCCCCCGTTGTACTCC 22
|||||

Db      786 GTGGAAGCCCCCGTTGTACTCC 765

RESULT 15
US-10-250-823-11/c
; Sequence 11, Application US/10250823
; Publication No. US20040072265A1
; GENERAL INFORMATION:
; APPLICANT: KEIO UNIVERSITY
; TITLE OF INVENTION: Human Brain Tumor Antigens and method for preparation
; FILE REFERENCE: P00000114
; CURRENT APPLICATION NUMBER: US/10/250,823
; CURRENT FILING DATE: 2003-07-07
; PRIOR APPLICATION NUMBER: JP 2001-001965
; PRIOR FILING DATE: 2001-01-09
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 3847
; TYPE: DNA
; ORGANISM: Homo sapiens
```

Mon Jul 19 07:46:11 2004

```
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (211)..(2883)
US-10-250-823-11
Query Match      92.7%; Score 20.4; DB 12; Length 3847;
Best Local Similarity 95.5%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 GTGGAAGCCGCGTTGTACTCC 22
         ||||| ||||| ||||| |||||
Db      30 GTGGAAGCCGCGTTGTACTCC 9
```

Search completed: July 18, 2004, 14:47:32
Job time : 39.5902 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 11:55:38 ; Search time 272.872 Seconds
(without alignments)
2407.604 Million cell updates/sec

Title: US-10-051-769-6
Perfect score: 22
Sequence: 1 gtggaagccgcgtgtactcc 22

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estin.*
- 4: em_estm.*
- 5: em_estov.*
- 6: em_estpl.*
- 7: em_estro.*
- 8: em_htc.*
- 9: gb_est1.*
- 10: gb_est2.*
- 11: gb_htc.*
- 12: gb_est3.*
- 13: gb_est4.*
- 14: gb_est5.*
- 15: em_estfun.*
- 16: em_estom.*
- 17: em_gss_hum.*
- 18: em_gss_inv.*
- 19: em_gss_pln.*
- 20: em_gss_vit.*
- 21: em_gss_fun.*
- 22: em_gss_mam.*
- 23: em_gss_mus.*
- 24: em_gss_pro.*
- 25: em_gss_rod.*
- 26: em_gss_phg.*
- 27: em_gss_vrl.*
- 28: gb_gss1.*
- 29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	22	100.0	730	BU449181	BU449181 603767430
C 2	22	100.0	898	13	BU453045
C 3	22	100.0	1145	28	CC290954
C 4	20.4	92.7	437	13	BY276266

C 5	20.4	92.7	458	10	BB638573
C 6	20.4	92.7	464	10	BB655451
C 7	20.4	92.7	583	10	BF330450
C 8	20.4	92.7	702	13	BU706646
C 9	20.4	92.7	738	29	CC567140
C 10	20.4	92.7	930	13	BQ960960
C 11	20.4	92.7	958	29	CNS04CSE
C 12	20.4	92.7	989	13	BX341552
C 13	20.4	92.7	2052	11	AK041847
C 14	20.4	92.7	3069	11	HC033019
C 15	19.4	88.2	399	11	AY232234
C 16	19.4	88.2	723	9	AW013379
C 17	19.4	88.2	836	13	BQ705427
C 18	18.8	85.5	456	29	CNS04KKQ
C 19	18.8	85.5	732	29	CNS04SQ4
C 20	18.8	85.5	790	28	AQ292068
C 21	18.8	85.5	795	29	CNS03XVR
C 22	18.8	85.5	933	29	CNS033H6
C 23	17.8	80.9	377	13	BQ500868
C 24	17.8	80.9	505	9	AA553509
C 25	17.8	80.9	651	14	CA350705
C 26	17.8	80.9	854	9	AU140984
C 27	17.8	80.9	1031	29	CNS02AYS
C 28	17.4	79.1	564	14	CD346734
C 29	17.2	78.2	305	13	C20150
C 30	17.2	78.2	363	13	BU430443
C 31	17.2	78.2	379	14	M79619
C 32	17.2	78.2	386	10	BF773382
C 33	17.2	78.2	410	14	CB804084
C 34	17.2	78.2	505	29	CE190332
C 35	17.2	78.2	507	29	PI79R
C 36	17.2	78.2	512	29	CG501015
C 37	17.2	78.2	517	14	CF650940
C 38	17.2	78.2	546	29	CC842724
C 39	17.2	78.2	551	14	W67915
C 40	17.2	78.2	560	12	BG19267
C 41	17.2	78.2	587	14	M79620
C 42	17.2	78.2	616	29	CC628362
C 43	17.2	78.2	638	14	CB017938
C 44	17.2	78.2	639	28	BZ327334
C 45	17.2	78.2	660	12	BM017266

ALIGNMENTS

RESULT 1	BU449181/c	BU449181	603767430F1	CSEQRBNI3	Gallus gallus	CDNA clone	ChEBST69518 5', mRNA	linear	EST 29-NOV-2002
LOCUS	BU449181	603767430F1	CSEQRBNI3	Gallus gallus	CDNA clone	ChEBST69518 5', mRNA	linear	EST 29-NOV-2002	
DEFINITION	BU449181	603767430F1	CSEQRBNI3	Gallus gallus	CDNA clone	ChEBST69518 5', mRNA	linear	EST 29-NOV-2002	
ACCESSION	BU449181	603767430F1	CSEQRBNI3	Gallus gallus	CDNA clone	ChEBST69518 5', mRNA	linear	EST 29-NOV-2002	
VERSION	BU449181.1	603767430F1	CSEQRBNI3	Gallus gallus	CDNA clone	ChEBST69518 5', mRNA	linear	EST 29-NOV-2002	
KEYWORDS	EST.	BU449181.1	603767430F1	CSEQRBNI3	Gallus gallus	CDNA clone	ChEBST69518 5', mRNA	linear	EST 29-NOV-2002
SOURCE	Gallus gallus (chicken)	BU449181	603767430F1	CSEQRBNI3	Gallus gallus	CDNA clone	ChEBST69518 5', mRNA	linear	EST 29-NOV-2002
ORGANISM	Gallus gallus	BU449181	603767430F1	CSEQRBNI3	Gallus gallus	CDNA clone	ChEBST69518 5', mRNA	linear	EST 29-NOV-2002
REFERENCE	1 (bases 1 to 730)	BU449181	603767430F1	CSEQRBNI3	Gallus gallus	CDNA clone	ChEBST69518 5', mRNA	linear	EST 29-NOV-2002
AUTHORS	Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Pong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.	BU449181	603767430F1	CSEQRBNI3	Gallus gallus	CDNA clone	ChEBST69518 5', mRNA	linear	EST 29-NOV-2002
TITLE	A Comprehensive Collection of Chicken cDNAs	BU449181	603767430F1	CSEQRBNI3	Gallus gallus	CDNA clone	ChEBST69518 5', mRNA	linear	EST 29-NOV-2002
JOURNAL	Curr. Biol. 12 (22), 1965-1969 (2002)	BU449181	603767430F1	CSEQRBNI3	Gallus gallus	CDNA clone	ChEBST69518 5', mRNA	linear	EST 29-NOV-2002
MEDLINE	22335534	BU449181	603767430F1	CSEQRBNI3	Gallus gallus	CDNA clone	ChEBST69518 5', mRNA	linear	EST 29-NOV-2002
PUBMED	12445392	BU449181	603767430F1	CSEQRBNI3	Gallus gallus	CDNA clone	ChEBST69518 5', mRNA	linear	EST 29-NOV-2002
COMMENT	Contact: Simon Hubbard Department of Biomolecular Sciences University of Manchester Institute of Science and Technology (UMIST) PO Box 88, Manchester, M60 1QD, UK Tel: 01612008930 Fax: 01612360409	BU449181	603767430F1	CSEQRBNI3	Gallus gallus	CDNA clone	ChEBST69518 5', mRNA	linear	EST 29-NOV-2002

Email: Simon.Hubbard@umist.ac.uk.

FEATURES

source

```

1. 730
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Layer"
/db_xref="taxon:9031"
/clone="ChEST69518"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQORN13"
/note="Organ: ovary; Vector: pBluescript II KS(+); Site_1:
EcoRI; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunted, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996) 791, except that a significantly longer
reannealing hybridization was used."

```

ORIGIN

```

Query Match      100.0%; Score 22; DB 13; Length 730;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGAGCGCCGCTTGTACTCC 22
    |||||
Db 376 GTGGAGCGCCGCTTGTACTCC 355

```

```

RESULT 2
BU453045/c
LOCUS      BU453045      898 bp  mRNA  linear  EST 29-NOV-2002
DEFINITION 603219082F1 CSEQORN14 Gallus gallus cDNA clone ChEST210c22 5', mRNA
sequence.
ACCESSION  BU453045
VERSION    BU453045.1  GI:259422356
KEYWORDS  EST.
SOURCE     Gallus gallus (chicken)
ORGANISM   Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 898)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
12445392
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1. 898
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Layer"
/db_xref="taxon:9031"
/clone="ChEST210c22"
/sex="Female"

```

FEATURES

source

```

1. 898
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Layer"
/db_xref="taxon:9031"
/clone="ChEST210c22"
/sex="Female"

```

```

/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQORN14"
/note="Organ: ovary; Vector: pBluescript II KS(+); Site_1:
EcoRI; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunted, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996) 791, except that a significantly longer
reannealing hybridization was used."

```

ORIGIN

```

Query Match      100.0%; Score 22; DB 13; Length 898;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGAGCGCCGCTTGTACTCC 22
    |||||
Db 376 GTGGAGCGCCGCTTGTACTCC 355

```

RESULT 3

```

CC290954
LOCUS      CC290954      1145 bp  DNA  linear  GSS 13-MAY-2003
DEFINITION CH261-100B2_Sp6.1 CH261 Gallus gallus genomic clone CH261-100B2,
genomic survey sequence.
ACCESSION  CC290954
VERSION    CC290954.1  GI:30662395
KEYWORDS  GSS.
SOURCE     Gallus gallus (chicken)
ORGANISM   Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 1145)
Kremitzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J.,
Warren, W., Graves, T., Mardis, E. and Wilson, R.
Gallus gallus BAC End Reads
Unpublished (2003)
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 182000 Std Error: 0.00
Seq primer: Sp6 ATTATGATGACACTATAG
Class: BAC ends
High quality sequence start: 28
High quality sequence stop: 628.
Location/Qualifiers
1. 1145
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="CH261-100B2"
/sex="female"
/cell_line="UCD001, inbred 256"
/clone_lib="CH261"
/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CH261 Female Chicken library - for library and clone
ordering information: http://www.chori.org/bacpac"

```

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

```

1. 1145
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="CH261-100B2"
/sex="female"
/cell_line="UCD001, inbred 256"
/clone_lib="CH261"
/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CH261 Female Chicken library - for library and clone
ordering information: http://www.chori.org/bacpac"

```

ORIGIN

```

Query Match      100.0%; Score 22; DB 28; Length 1145;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Michela Fagioli and Takao K. Hensch (Laboratory for Neuronal Circuit Development Brain Science Institute RIKEN 2-1 Hirose, Wako-shi, Saitama 351-0198 Japan) whose assistance we gratefully acknowledge. Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

1. .437
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K430318E15"
/tissue_type="visual cortex"
/clone_lib="RIKEN full-length enriched, visual cortex"

ORIGIN
Query Match 92.7%; Score 20.4; DB 13; Length 437;
Best Local Similarity 95.5%; Pred. No. 5.8e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGGAAGCCGCGTGTACTCC 22
|||||
Db 324 GTGGAAGCCGCGTGTACTCC 303
|||||

RESULT 5
BB638573/c
LOCUS
DEFINITION
BB638573 RIKEN full-length enriched, 3 days neonate thymus Mus musculus cDNA clone A630041P07 5', mRNA sequence.
BB638573
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 GTGGAAGCCGCGTGTACTCC 22
|||||
527 GTGGAAGCCGCGTGTACTCC 548
|||||

BY276266 437 bp mRNA linear EST 11-DEC-2002
BY276266 RIKEN full-length enriched, visual cortex Mus musculus
cDNA clone K430318E15 5', mRNA sequence.
BY276266
BY276266.1 GI:26466603
EST.
Mus musculus (house mouse)
Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 437)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikiado, I., Osato, N., Saito, R., Suzuki, H., Yamahata, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Rume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fietcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Perte, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenawa, C., Taylor, M. S., Teasdale, R. D., Tomita, M., Varardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
22354683
12466851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome.res@gs.riken.go.jp
URL: <http://genome.gsc.riken.go.jp/>
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res.

and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Yamanaka, I., Kiyosawa, H., Kondo, S., Saito, T., Shinagawa, A., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., Arakawa, T., Ishii, Y. and Hayashizaki, Y.
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct. Funct. Genomics 2 pre, L72-L86 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
e mouse tissues.

FEATURES source
Location/Qualifiers
1..458
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="A630041P07"
/tissue_type="thymus"
/dev_stage="3 days neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 3 days neonate thymus"
/note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5',
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATTCGAGTTAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

Query Match 92.7%; Score 20.4; DB 10; Length 458;
Best Local Similarity 95.5%; Pred. No. 5.8e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GTGAAGCCGCGTGTACTCC 22
Db 249 GTGGAACCGCGTGTACTCC 228
BB655451 464 bp mRNA linear EST 26-OCT-2001
BB655451 RIKEN full-length enriched, 9 days embryo Mus musculus
cDNA clone D030060C17 5', mRNA sequence.
BB655451
BB655451.1 GI:16489279
EST.
Mus musculus (house mouse)

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 464);
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,

Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsuura, S., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
e mouse tissues.

FEATURES source

Location/Qualifiers
1..464
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="D030060C17"
/dev_stage="9 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 9 days embryo"
/note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5',
GAGAGAGAGCGCCGCAACTCGAGTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5',
GAGAGAGATTCGAGTTAATTAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

Query Match 92.7%; Score 20.4; DB 10; Length 464;
Best Local Similarity 95.5%; Pred. No. 5.8e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GTGAAGCCGCGTGTACTCC 22
Db 75 GTGGAACCGCGTGTACTCC 54

```

ORGANISM      Mus musculus
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS       Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE         1 (bases 1 to 702)
JOURNAL       NIH-MGC http://mgc.nci.nih.gov/.
COMMENT       National Institutes of Health, Mammalian Gene Collection (MGC)
              Unpublished (1999)
              Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-remail.nih.gov
              Tissue Procurement: Dr. Jim Lin, University of Iowa
              cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
              cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
              cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa
              Clone Distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              This clone was contributed by the Brain Molecular Anatomy Project
              (BMAP)
Seq primer: pyX-5.
Location/Qualifiers
    1..702
        /organism="Mus musculus"
        /mol_type="mRNA"
        /strain="C57BL/6"
        /db_xref="taxon:10090"
        /clone="IMAGE:6409978"
        /tissue_type="whole brain"
        /dev_stage="embryo 12.5dpc"
        /lab_host="DH10B (T1 phage resistant)"
        /clone_lib="NIH-BMAP_FOQ"
        /note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;
            Site 2: Not I; The library was constructed according
            Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
            1996. Denatured RNA was size fractionated on a 1% agarose
            gel. First strand cDNA synthesis was primed with oligo-dT
            primer containing a Not I site. Double strand cDNA was
            size selected according to mRNA size fraction, ligated
            with EcoR I adaptor, digested with NotI and then cloned
            directionally into pYX-Asc vector. The library tag
            sequence located between the Not I site and the polyA tail
            is TCAGACGACC. This library was created for the University
            Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the
            Developing Mouse Nervous System', supported by National
            Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
            program coordinator."
ORIGIN
Query Match          92.7%; Score 20.4; DB 13; Length 702;
Best Local Similarity 95.5%; Pred. No. 6.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1   GTGGAGCGCCGCTGTACTCC 22
        ||||| ||||||| |||||
Db      553 GTGGAACCGCGCTGTACTCC 532

RESULT 9
LOCUS    CC567140/c
DEFINITION CH240_441017.TARBAcl3P2 CHORI-240 Bos taurus genomic clone
ACCESSION CC567140
VERSION   CC567140.1 GI:31900496
KEYWORDS  GSS.
SOURCE   Bos taurus (cow)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;
          Bovidae; Bovinae; Bos.
          1 (bases 1 to 738)
          Holt,R., Stott,J., Yang,G., Barber,S., Smalhus,D., Prabhu,A.-L.,
          Tsai,M., Cloutier,A., Lee,D., Girn,N., Olson,T., Mayo,M.,
          Butterfield,Y., Kirkpatrick,R., Liu,J., Guin,R., Chan,A., Chiu,R.,

```

Mathewson, C., Wye, N., Masson, A., Brown-John, M., Jones, S., Schein, J., Marra, M., de Jong, P., Keele, J. W. and Kappes, S. M. Bovine BAC End Sequences from Library CHORI-240, PLATES 399 to 478. Unpublished (2003)

Other GSSs: CH240.441017.T7

Contact: Rob Holt

Sequencing

The British Columbia Cancer Agency Genome Science Centre
600 W. 10th Ave, Vancouver, British Columbia, Canada V5Z 4E6
Tel: 604-877-6085
Fax: 604-877-6276
Email: rholt@bcgsc.ca

Clones are derived from the bovine BAC library CHORI-240 (<http://www.chori.org/bacpac/bovine240.htm>). For BAC library availability, please contact Pieter de Jong (pdjong@mail.choi.org). Clones may be purchased from BACPAC Resources (<http://www.chori.org/bacpac/ordering-information.htm>). This work was undertaken as part of the International Bovine BAC Mapping Consortium (IBBMC) by CSIRO Livestock Industries, Australia and the British Columbia Genome Science Centre, Canada.

Plate: 441 row: 0 column: 17
Seq primer: SP6
Class: BAC ends.

FEATURES source

Location/Qualifiers
1. .738
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="breed: Hereford"
/db_xref="taxon:9913"
/clone="CH240_441017"
/sex="Male"
/cell_type="Blood"
/clone_lib="CHORI-240"
/note="Vector: pTARBAC1.3; Site 1: Mbol; Site 2: Mbol; Hereford bull 1L Domino 99375; CHORI-240 Bovine BAC library (Male) produced by Pieter de Jong"

ORIGIN

Query Match 92.7%; Score 20.4; DB 29; Length 738;
Best Local Similarity 95.5%; Pred. No. 6.3e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGGAAGCCGCGTTGTACTCC 22
|||||
Db 589 GTGGAAGCCGCGTTGTACTCC 568

RESULT 10
BQ960960/c
LOCUS
DEFINITION BQ960960.1 NC1_CGAP Mam2 Mus musculus cDNA clone
IMAGE:6439792 5', mRNA sequence.
ACCESSION BQ960960
VERSION BQ960960.1 GI:22376438
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 930)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLNL13960 row: f column: 17

High quality sequence stop: 603.
Location/Qualifiers
1. .930
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:6439792"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP Mam2"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: Sall; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

ORIGIN

Query Match 92.7%; Score 20.4; DB 13; Length 930;
Best Local Similarity 95.5%; Pred. No. 6.6e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGGAAGCCGCGTTGTACTCC 22
|||||
Db 578 GTGGAAGCCGCGTTGTACTCC 557

RESULT 11 CNS04CSE/c LOCUS

DEFINITION CNS04CSE 958 bp DNA linear GSS 01-SEP-2000
Tetraodon nigroviridis genome survey sequence 17 end of clone 101121 of library G from Tetraodon nigroviridis, genomic survey sequence.
AL284855
VERSION AL284855.1 GI:8023256
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.

REFERENCE

AUTHORS Roest Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.
Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)

JOURNAL
MEDLINE
PUBMED
REFERENCE

AUTHORS Roest Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)

TITLE

JOURNAL
MEDLINE
PUBMED
REFERENCE

AUTHORS Roest Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE

AUTHORS Roest Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)

COMMENT

This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES

source
Location/Qualifiers
1. .958
/organism="Tetraodon nigroviridis"

```

/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="101121"
/clone_lib="G"
/note="Genoscope sequence ID : COBG101AE11LP1-end : T7"

ORIGIN
Query Match          92.7%; Score 20.4; DB 29; Length 958;
Best Local Similarity 95.5%; Pred. No. 6.6e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGGAAGCGCGGTGTACTCC 22
    |||||
Db 106 GTGGAAGCGCGGTGTACTCC 85

RESULT 12
LOCUS BX341552/c
DEFINITION BX341552 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
cDNA clone CS0DK002YA12 5-PRIME, mRNA sequence.
ACCESSION BX341552
VERSION BX341552.1 GI:30339998
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 989)
AUTHORS Li W.B., Gruber C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
COMMENT Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5483.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DK002BA06QP1&cluster=5483.r. Contact :
Feng liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DK002BA06QP1.
FEATURES
    source
    Location/Qualifiers
        1..989
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0DK002YA12"
            /cell_type="HELA CELLS COT 25-NORMALIZED"
            /cell_lines="HELA"
            /clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
            /note="1st strand cDNA was primed with a NotI-oligo(dT)
            primer. Five prime end enriched, double-strand cDNA was
            digested with Not I and cloned into the Not I and EcoR V
            sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match          92.7%; Score 20.4; DB 13; Length 989;
Best Local Similarity 95.5%; Pred. No. 6.7e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGGAAGCGCGGTGTACTCC 22
    |||||
Db 876 GTGGAAGCGCGGTGTACTCC 855

RESULT 13
LOCUS AK041847/c
DEFINITION Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
library, clone:A630041P07 product:hypothetical PDZ domain (also
known as DHR or GLGF) containing protein, full insert sequence.
AK041847
VERSION AK041847.1 GI:26088683
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 9279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,I., Harada,A.,
Yanamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
PUBMED 11076861
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
PUBMED 12052161
REFERENCE 6
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Atakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,T., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koyama,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toyota,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan [E-mail:genome-res@gs.riken.go.jp,
URL:http://genome-gsc.riken.go.jp/, Tel.81-45-503-9222,
Fax:81-45-503-9216]
COMMENT cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics

```

Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MSC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.
Please visit our web site for further details.
URL: <http://genome.gsc.riken.go.jp/>
URL: <http://fantom.gsc.riken.go.jp/>.

FEATURES

source

1..2052
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:A630041P07"
/db_xref="MGI:2406859"
/db_xref="taxon:10090"
/clone="A630041P07"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="3 days neonate"
misc_feature
1..2052
/note="hypothetical PDZ domain (also known as DHR or GLGF) containing protein (InterPro|IPR001478, evidence: InterPro)"

ORIGIN

Query Match 92.7%; Score 20.4; DB 11; Length 2052;
Best Local Similarity 95.5%; Pred. No. 7.5e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GTGGAAGCCCGCTTGTACTCC 22
|||||
DB 253 GTGGAAGCCCGCTTGTACTCC 232

RESULT 14

BC033019/c

LOCUS

DEFINITION Homo sapiens hypothetical protein LOC118987, mRNA (cDNA clone IMAGE:4821752), containing frame-shift errors.

ACCESSION BC033019

VERSION BC033019.1

KEYWORDS GI:21542573

SOURCE HTc.

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 3069)

Strausberg,R.L., Feigold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Schetz,T.E., Brownstein,M.J., Ussid,T.B., Toshiyuki,S., Carninci,P., Frange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullaly,S.J., Bosak,S.A., McSwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahy,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S., Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalls,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

MEDLINE 12477932

PUBMED 2 (bases 1 to 3069)

REFERENCE Strausberg,R.

Direct Submission

TITLE

Submitted (21-JUN-2002) National Institutes of Health, Mammalian

REMARK

COMMENT

Location/Qualifiers

1..3069

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4821752"

/tissue_type="Testis"

/clone_lib="NIH MGC 97"

/lab_host="DH10B"

/note="Vector: pBluescript"

Query Match 92.7%; Score 20.4; DB 11; Length 3069;

Best Local Similarity 95.5%; Pred. No. 8.1e+02;

Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGGAAGCCCGCTTGTACTCC 22

|||||

DB 853 GTGGAAGCCCGCTTGTACTCC 832

RESULT 15

AY232234

LOCUS

DEFINITION Drosophila yakuba clone yak-ad_CG2254 mRNA sequence.

ACCESSION AY232234

VERSION AY232234.1

KEYWORDS GI:38048708

SOURCE HTc.

ORGANISM Drosophila yakuba

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 399)

Domazet-Lozo,T. and Tautz,D.

An evolutionary analysis of orphan genes in Drosophila

Genome Res. 13 (10), 2213-2219 (2003)

22887302

PUBMED 14525923

REFERENCE 2 (bases 1 to 399)

Domazet-Lozo,T. and Tautz,D.

Direct Submission

TITLE

Submitted (07-FEB-2003) Evolution Genetics, Institute for Genetics,

Weyertal 121, Cologne 50931, Germany

Location/Qualifiers

1..399

/organism="Drosophila yakuba"

/mol_type="mRNA"

/db_xref="taxon:7245"

/clone="yak-ad_CG2254"

<1..>399

CDS

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadansystemsbiology.org
Anup Madan, Jessica Fahey, Brin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 33 Row: b Column: 4
This clone was selected for full length sequencing because it passed the following selection criteria: GenomScan gene prediction
This clone has the following problem: frame shifted.

FEATURES

source

1..3069
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4821752"
/tissue_type="Testis"
/clone_lib="NIH MGC 97"
/lab_host="DH10B"
/note="Vector: pBluescript"

ORIGIN

Query Match 92.7%; Score 20.4; DB 11; Length 3069;
Best Local Similarity 95.5%; Pred. No. 8.1e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GTGGAAGCCCGCTTGTACTCC 22
|||||
DB 853 GTGGAAGCCCGCTTGTACTCC 832

RESULT 15

AY232234

LOCUS

DEFINITION Drosophila yakuba clone yak-ad_CG2254 mRNA sequence.

ACCESSION AY232234

VERSION AY232234.1

KEYWORDS GI:38048708

SOURCE HTc.

ORGANISM Drosophila yakuba

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 399)

Domazet-Lozo,T. and Tautz,D.

An evolutionary analysis of orphan genes in Drosophila

Genome Res. 13 (10), 2213-2219 (2003)

22887302

PUBMED 14525923

REFERENCE 2 (bases 1 to 399)

Domazet-Lozo,T. and Tautz,D.

Direct Submission

TITLE

Submitted (07-FEB-2003) Evolution Genetics, Institute for Genetics,

Weyertal 121, Cologne 50931, Germany

Location/Qualifiers

1..399

/organism="Drosophila yakuba"

/mol_type="mRNA"

/db_xref="taxon:7245"

/clone="yak-ad_CG2254"

<1..>399

CDS

/codon_start=1
/product="similar to Drosophila melanogaster CG2254"
/protein_id="AAR10257.1"
/db_xref="GI:38048709"
/translation="MSKVTSPAGNNANSNDIYNIVLVVDIVMLIVKFWIAVVEAAVG
LFRPAPLDDVSGKVVLTGTGHGKGQALQYAKLGAILLCWDVNEQTNQTVKEIKS
SGGKAFGYVCNVTKREELIELAQVKRKHGF"

ORIGIN

Query Match 88.2%; Score 19.4; DB 11; Length 399;
Best Local Similarity 95.2%; Pred. No. 1.4e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GTGGAAGCCGCGTGTGACTC 21
Db 115 GTGGAAGCCGCGTGTGACTC 135

Search completed: July 18, 2004, 14:36:39
Job time : 276.872 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 12:36:30 ; Search time 457.82 Seconds
(without alignments)
2781.271 Million cell updates/sec

Title: US-10-051-769-2
Perfect score: 261
Sequence: 1 gatcaaggtgagttcgagg.....cacttgcccatcgactgga 261

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3190992 seqs, 2439311697 residues

Total number of hits satisfying chosen parameters: 6381984

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	261	100.0	261	15	US-10-051-769-2	Sequence 2, Appli
2	261	100.0	261	15	US-10-224-624-2	Sequence 2, Appli
3	261	100.0	873	15	US-10-051-769-3	Sequence 3, Appli
4	261	100.0	873	15	US-10-224-624-3	Sequence 3, Appli
5	261	100.0	960	15	US-10-051-769-1	Sequence 1, Appli
6	261	100.0	960	15	US-10-224-624-1	Sequence 1, Appli
7	261	100.0	2639	16	US-10-104-047-684	Sequence 684, App
8	261	100.0	3465	15	US-10-224-624-9	Sequence 9, Appli
9	261	100.0	3649	13	US-10-112-944-63	Sequence 63, Appli
10	261	100.0	3932	15	US-10-224-624-7	Sequence 7, Appli
11	261	100.0	4470	13	US-10-276-774-950	Sequence 950, App
12	261	100.0	4801	12	US-10-336-603A-25	Sequence 25, Appli
13	208.2	79.8	263	15	US-10-051-769-4	Sequence 4, Appli
14	208.2	79.8	263	15	US-10-224-624-4	Sequence 4, Appli

15	51.2	19.6	11058	15	US-10-156-761-3629	Sequence 3629, Ap
16	51.2	19.6	9025608	15	US-10-156-761-1	Sequence 1, Appli
17	50.4	19.3	848	13	US-10-425-114-16261	Sequence 16261, A
18	49.8	19.1	3243	15	US-10-156-761-2685	Sequence 2685, A
19	49.8	19.1	9025608	15	US-10-156-761-1	Sequence 1, Appli
20	49.4	18.9	2499	15	US-10-156-761-6721	Sequence 6721, Ap
21	47.8	18.3	2710	17	US-10-250-613-33	Sequence 33, Appli
22	47.4	18.2	873	15	US-10-156-761-3317	Sequence 3317, Ap
23	47.4	18.2	1590	15	US-10-156-761-5921	Sequence 5921, Ap
24	47.2	18.1	1271	17	US-10-437-963-101203	Sequence 101203, A
25	47	18.0	552	17	US-10-437-963-87461	Sequence 87461, A
26	47	18.0	1008	17	US-10-437-963-45694	Sequence 45694, A
27	47	18.0	3847	12	US-10-250-823-11	Sequence 11, Appli
28	47	18.0	10232	13	US-10-282-122A-25487	Sequence 25487, A
29	46.8	17.9	1392	15	US-10-156-761-7375	Sequence 7375, Ap
30	46.8	17.9	1721	17	US-10-437-963-78212	Sequence 78212, A
31	46	17.6	809	17	US-10-437-963-12363	Sequence 12363, A
32	46	17.6	825	15	US-10-156-761-7318	Sequence 7318, Ap
33	46	17.6	1431	12	US-09-758-759-136	Sequence 136, App
34	46	17.6	1494	15	US-10-156-761-2713	Sequence 2713, Ap
35	46	17.6	109519	12	US-09-758-759-1	Sequence 1, Appli
36	45.8	17.5	985	17	US-10-437-963-29072	Sequence 29072, A
37	45.4	17.4	927	15	US-10-156-761-3926	Sequence 3926, Ap
38	45.4	17.4	975	15	US-10-156-761-4570	Sequence 4570, Ap
39	45.4	17.4	1224	17	US-10-437-963-56422	Sequence 56422, A
40	45.4	17.4	1335	15	US-10-156-761-1454	Sequence 1454, Ap
41	45.2	17.3	659	15	US-10-029-386-20268	Sequence 20268, A
42	45.2	17.3	2973	15	US-10-156-761-2660	Sequence 2660, Ap
43	45.2	17.3	4320	16	US-10-440-464-139	Sequence 139, App
44	45	17.2	1163	13	US-10-425-114-22074	Sequence 22074, A
45	45	17.2	1277	13	US-10-425-114-32433	Sequence 32433, A

ALIGNMENTS

RESULT 1
US-10-051-769-2
; Sequence 2, Application US/10051769
; Publication No. US20030044811A1
; GENERAL INFORMATION:
; APPLICANT: McKINNON, Randy D.
; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION
; FILE REFERENCE: 268/260 (RWJ-00-37)
; CURRENT APPLICATION NUMBER: US/10/051,769
; CURRENT FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,160
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 261
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-051-769-2

Query Match 100.0%; Score 261; DB 15; Length 261;
Best Local Similarity 100.0%; Pred. No. 3.5e-58;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GATCAAGGTGAGTTTCGAGGAGCTCTGAGACCAAGACGGCGCGGCGCTCTGAGGG	60
Db	1	GATCAAGGTGAGTTTCGAGGAGCTCTGAGACCAAGACGGCGCGGCGCTCTGAGGG	60
Qy	61	GCTGACGCTCGCGGACGTTTCTTGGCGGAGACGCTTTCATCAAGACCATCCGGCT	120
Db	61	GCTGACGCTCGCGGACGTTTCTTGGCGGAGACGCTTTCATCAAGACCATCCGGCT	120
Qy	121	CCTGCGGCGAGTCGTGCCCTCGGCCACCGGGAGCCCGATGTCCTGAAGGGAGCGCT	180
Db	121	CCTGCGGCGAGTCGTGCCCTCGGCCACCGGGAGCCCGATGTCCTGAAGGGAGCGCT	180
Qy	181	GCCTGCGGCGCTCCCGGAGGAGCTGGCTTCGAGCGGAGGTGGAGTACACGGGGGCTT	240

Db 181 GCCCGCCGCTGCGCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGAGTACAAACGGGGGCTT 240
QY 241 CCACCTGGCCCATCGACGTGGA 261
Db 241 CCACCTGGCCCATCGACGTGGA 261

RESULT 2

US-10-224-624-2
; Sequence 2, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein Gliten
; FILE REFERENCE: 54704.8059.US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 261
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-224-624-2

Query Match 100.0%; Score 261; DB 15; Length 261;
Best Local Similarity 100.0%; Pred. No. 3.5e-58;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCGGCGCTGCTGGAGGG 60
Db 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCGGCGCTGCTGGAGGG 60
QY 61 GCTGAGCCTCGCGGAGCTGTCTCTGGCGGAGAGCGGTGCTTCATCAAGACCATCCGGCT 120
Db 61 GCTGAGCCTCGCGGAGCTGTCTCTGGCGGAGAGCGGTGCTTCATCAAGACCATCCGGCT 120
QY 121 CGTGGCGCCAGTGTGCTCCCTCGCCACCGCGGAGCCGATGGCCCTGAAGGGAGGCGCT 180
Db 121 CGTGGCGCCAGTGTGCTCCCTCGCCACCGCGGAGCCGATGGCCCTGAAGGGAGGCGCT 180
QY 181 GCCCGCCGCTGCGCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGAGTACAAACGGGGGCTT 240
Db 181 GCCCGCCGCTGCGCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGAGTACAAACGGGGGCTT 240
QY 241 CCACCTGGCCCATCGACGTGGA 261
Db 241 CCACCTGGCCCATCGACGTGGA 261

RESULT 3

US-10-051-769-3
; Sequence 3, Application US/10051769
; Publication No. US20030044811A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randy D.
; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION
; FILE REFERENCE: 268/260 (RMJ-00-37)
; CURRENT APPLICATION NUMBER: US/10/051,769
; CURRENT FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,160
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 873
; TYPE: DNA
; ORGANISM: Homo Sapiens

US-10-051-769-3

Query Match 100.0%; Score 261; DB 15; Length 873;
Best Local Similarity 100.0%; Pred. No. 3e-58;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCGGCGCTGCTGGAGGG 60
Db 366 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCGGCGCTGCTGGAGGG 425
QY 61 GCTGAGCCTCGCGGAGCTGTCTCTGGCGGAGAGCGGTGCTTCATCAAGACCATCCGGCT 120
Db 426 GCTGAGCCTCGCGGAGCTGTCTCTGGCGGAGAGCGGTGCTTCATCAAGACCATCCGGCT 485
QY 121 CGTGGCGCCAGTGTGCTCCCTCGCCACCGCGGAGCCGATGGCCCTGAAGGGAGGCGCT 180
Db 486 CGTGGCGCCAGTGTGCTCCCTCGCCACCGCGGAGCCGATGGCCCTGAAGGGAGGCGCT 545
QY 181 GCCCGCCGCTGCGCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGAGTACAAACGGGGGCTT 240
Db 546 GCCCGCCGCTGCGCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGAGTACAAACGGGGGCTT 605
QY 241 CCACCTGGCCCATCGACGTGGA 261
Db 606 CCACCTGGCCCATCGACGTGGA 626

RESULT 4

US-10-224-624-3
; Sequence 3, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein Gliten
; FILE REFERENCE: 54704.8059.US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 873
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-224-624-3

Query Match 100.0%; Score 261; DB 15; Length 873;
Best Local Similarity 100.0%; Pred. No. 3e-58;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCGGCGCTGCTGGAGGG 60
Db 366 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCGGCGCTGCTGGAGGG 425
QY 61 GCTGAGCCTCGCGGAGCTGTCTCTGGCGGAGAGCGGTGCTTCATCAAGACCATCCGGCT 120
Db 426 GCTGAGCCTCGCGGAGCTGTCTCTGGCGGAGAGCGGTGCTTCATCAAGACCATCCGGCT 485
QY 121 CGTGGCGCCAGTGTGCTCCCTCGCCACCGCGGAGCCGATGGCCCTGAAGGGAGGCGCT 180
Db 486 CGTGGCGCCAGTGTGCTCCCTCGCCACCGCGGAGCCGATGGCCCTGAAGGGAGGCGCT 545
QY 181 GCCCGCCGCTGCGCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGAGTACAAACGGGGGCTT 240
Db 546 GCCCGCCGCTGCGCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGAGTACAAACGGGGGCTT 605
QY 241 CCACCTGGCCCATCGACGTGGA 261
Db 606 CCACCTGGCCCATCGACGTGGA 626

```

RESULT 5
US-10-051-769-1
; Sequence 1, Application US/10051769
; Publication No. US2003004481A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randy D.
; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION
; FILE REFERENCE: 268/260 (RWJ-00-37)
; CURRENT APPLICATION NUMBER: US/10/051,769
; CURRENT FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,160
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Homo Sapiens
; US-10-051-769-1

Query Match      100.0%; Score 261; DB 15; Length 960;
Best Local Similarity 100.0%; Pred. No. 3e-58;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 60
Db      396  GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 455

QY      61  GCTGAGCTCGGGACGTGTTCTTGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
Db      456  GCTGAGCTCGGGACGTGTTCTTGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 515

QY      121  CGTGGCGCCAGTCGTGCCCTCGGCACCGGGAGCCCGATGGCCCTGAAGGGGAGCGCT 180
Db      516  CGTGGCGCCAGTCGTGCCCTCGGCACCGGGAGCCCGATGGCCCTGAAGGGGAGCGCT 575

QY      181  GCCCGCGCGCTGCCCCGAGGAGCTGGCCCTTCGAGGCGGAGGTGGAGTACAAACGGGGGCTT 240
Db      576  GCCCGCGCGCTGCCCCGAGGAGCTGGCCCTTCGAGGCGGAGGTGGAGTACAAACGGGGGCTT 635

QY      241  CCACCTGGCCATCGACGTGA 261
Db      636  CCACCTGGCCATCGACGTGA 656

;
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-104-047-684

Query Match      100.0%; Score 261; DB 16; Length 2639;
Best Local Similarity 100.0%; Pred. No. 2.7e-58;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 60
Db      579  GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 638

QY      61  GCTGAGCTCGGGACGTGTTCTTGGCGGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
Db      639  GCTGAGCTCGGGACGTGTTCTTGGCGGAGACGGTGCCTTCATCAAGACCATCCGGCT 698

QY      121  CGTGGCGCCAGTCGTGCCCTCGGCACCGGGAGCCCGATGGCCCTGAAGGGGAGCGCT 180
Db      699  CGTGGCGCCAGTCGTGCCCTCGGCACCGGGAGCCCGATGGCCCTGAAGGGGAGCGCT 758

QY      181  GCCCGCGCGCTGCCCCGAGGAGCTGGCCCTTCGAGGCGGAGGTGGAGTACAAACGGGGGCTT 240
Db      759  GCCCGCGCGCTGCCCCGAGGAGCTGGCCCTTCGAGGCGGAGGTGGAGTACAAACGGGGGCTT 818

QY      241  CCACCTGGCCATCGACGTGA 261
Db      819  CCACCTGGCCATCGACGTGA 839

;
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-104-047-684

Query Match      100.0%; Score 261; DB 15; Length 960;
Best Local Similarity 100.0%; Pred. No. 3e-58;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 60
Db      396  GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 455

QY      61  GCTGAGCTCGGGACGTGTTCTTGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
Db      456  GCTGAGCTCGGGACGTGTTCTTGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 515

QY      121  CGTGGCGCCAGTCGTGCCCTCGGCACCGGGAGCCCGATGGCCCTGAAGGGGAGCGCT 180
Db      516  CGTGGCGCCAGTCGTGCCCTCGGCACCGGGAGCCCGATGGCCCTGAAGGGGAGCGCT 575

QY      181  GCCCGCGCGCTGCCCCGAGGAGCTGGCCCTTCGAGGCGGAGGTGGAGTACAAACGGGGGCTT 240
Db      576  GCCCGCGCGCTGCCCCGAGGAGCTGGCCCTTCGAGGCGGAGGTGGAGTACAAACGGGGGCTT 635

QY      241  CCACCTGGCCATCGACGTGA 261
Db      636  CCACCTGGCCATCGACGTGA 656

;
; TYPE: DNA
; ORGANISM: Homo Sapiens
; US-10-051-769-1

Query Match      100.0%; Score 261; DB 15; Length 960;
Best Local Similarity 100.0%; Pred. No. 3e-58;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 60

```

```

Db      396  GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 455

QY      61  GCTGAGCTCGGGACGTGTTCTTGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
Db      456  GCTGAGCTCGGGACGTGTTCTTGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 515

QY      121  CGTGGCGCCAGTCGTGCCCTCGGCACCGGGAGCCCGATGGCCCTGAAGGGGAGCGCT 180
Db      516  CGTGGCGCCAGTCGTGCCCTCGGCACCGGGAGCCCGATGGCCCTGAAGGGGAGCGCT 575

QY      181  GCCCGCGCGCTGCCCCGAGGAGCTGGCCCTTCGAGGCGGAGGTGGAGTACAAACGGGGGCTT 240
Db      576  GCCCGCGCGCTGCCCCGAGGAGCTGGCCCTTCGAGGCGGAGGTGGAGTACAAACGGGGGCTT 635

QY      241  CCACCTGGCCATCGACGTGA 261
Db      636  CCACCTGGCCATCGACGTGA 656

;
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-104-047-684

Query Match      100.0%; Score 261; DB 16; Length 2639;
Best Local Similarity 100.0%; Pred. No. 2.7e-58;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 60
Db      579  GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 638

QY      61  GCTGAGCTCGGGACGTGTTCTTGGCGGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
Db      639  GCTGAGCTCGGGACGTGTTCTTGGCGGAGACGGTGCCTTCATCAAGACCATCCGGCT 698

QY      121  CGTGGCGCCAGTCGTGCCCTCGGCACCGGGAGCCCGATGGCCCTGAAGGGGAGCGCT 180
Db      699  CGTGGCGCCAGTCGTGCCCTCGGCACCGGGAGCCCGATGGCCCTGAAGGGGAGCGCT 758

QY      181  GCCCGCGCGCTGCCCCGAGGAGCTGGCCCTTCGAGGCGGAGGTGGAGTACAAACGGGGGCTT 240
Db      759  GCCCGCGCGCTGCCCCGAGGAGCTGGCCCTTCGAGGCGGAGGTGGAGTACAAACGGGGGCTT 818

QY      241  CCACCTGGCCATCGACGTGA 261
Db      819  CCACCTGGCCATCGACGTGA 839

;
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-104-047-684

Query Match      100.0%; Score 261; DB 15; Length 960;
Best Local Similarity 100.0%; Pred. No. 3e-58;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 60

```

```

RESULT 8
US-10-224-624-9
; Sequence 9, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GliTEN
; FILE REFERENCE: 54704.8059.US00

```

; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 3465
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-224-624-9

Query Match 100.0%; Score 261; DB 15; Length 3465;
Best Local Similarity 100.0%; Pred. No. 2.6e-58;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCGCTGCTGGAGGG 60
Db 366 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCGCTGCTGGAGGG 425
QY 61 GCTGAGCCTGCGGGAGCTGTTCTTGGCGAGAGCGTGCCTTCATCAAGACCATCCGGCT 120
Db 426 GCTGAGCCTGCGGGAGCTGTTCTTGGCGAGAGCGTGCCTTCATCAAGACCATCCGGCT 485
QY 121 CGTGGCGGCAGTCGTGCCCTCGGCCACCGGGGAGCCCGATGCCCTGAAGGGGAGCGCT 180
Db 486 CGTGGCGGCAGTCGTGCCCTCGGCCACCGGGGAGCCCGATGCCCTGAAGGGGAGCGCT 545
QY 181 GCCCGCGCGCTCCCGGAGAGCTGCGCTTCGAGGCGGAGGTGGAGTACAACGGGGGCTT 240
Db 546 GCCCGCGCGCTCCCGGAGAGCTGCGCTTCGAGGCGGAGGTGGAGTACAACGGGGGCTT 605
QY 241 CCACCTGGCCATCGACGTGGA 261
Db 606 CCACCTGGCCATCGACGTGGA 626

RESULT 9

US-10-112-944-63
; Sequence 63, Application US/10112944
; Publication No. US20040048249A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Gezhi
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Xue, Aidong J.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wehrman, Tom
; APPLICANT: Ghosh, Malabika
; APPLICANT: Wang, Dunrui
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Zhiwei
; TITLE OF INVENTION: No. US20040048249A1el Nucleic Acids and
; FILE REFERENCE: 805A
; CURRENT APPLICATION NUMBER: US/10/112,944
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: US 09/519,705
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/540,217
; PRIOR FILING DATE: 2000-03-31

; PRIOR APPLICATION NUMBER: US 09/552,929
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/577,408
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 924
; SOFTWARE: pt_FL_genes Version 5.0
; SEQ ID NO 63
; LENGTH: 3649
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3462)
US-10-112-944-63

Query Match 100.0%; Score 261; DB 13; Length 3649;
Best Local Similarity 100.0%; Pred. No. 2.6e-58;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCGCTGCTGGAGGG 60
Db 366 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCGCTGCTGGAGGG 425
QY 61 GCTGAGCCTGCGGGAGCTGTTCTTGGCGAGAGCGTGCCTTCATCAAGACCATCCGGCT 120
Db 426 GCTGAGCCTGCGGGAGCTGTTCTTGGCGAGAGCGTGCCTTCATCAAGACCATCCGGCT 485
QY 121 CGTGGCGGCAGTCGTGCCCTCGGCCACCGGGGAGCCCGATGCCCTGAAGGGGAGCGCT 180
Db 486 CGTGGCGGCAGTCGTGCCCTCGGCCACCGGGGAGCCCGATGCCCTGAAGGGGAGCGCT 545
QY 181 GCCCGCGCGCTCCCGGAGAGCTGCGCTTCGAGGCGGAGGTGGAGTACAACGGGGGCTT 240
Db 546 GCCCGCGCGCTCCCGGAGAGCTGCGCTTCGAGGCGGAGGTGGAGTACAACGGGGGCTT 605
QY 241 CCACCTGGCCATCGACGTGGA 261
Db 606 CCACCTGGCCATCGACGTGGA 626

RESULT 10

US-10-224-624-7
; Sequence 7, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: McKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GlITEN
; FILE REFERENCE: 54704.8059.US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 3832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (178)..(3639)
; OTHER INFORMATION:
US-10-224-624-7

Query Match 100.0%; Score 261; DB 15; Length 3832;
Best Local Similarity 100.0%; Pred. No. 2.6e-58;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCGCTGCTGGAGGG 60
Db 543 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCGCTGCTGGAGGG 602

```
QY 61 GCTGAGCCTCGGGACGTTCTCTGGCGGAGACGGTGGCCCTTCATCAAGACCATCCGGCT 120
Db 603 GCTGAGCCTCGGGACGTTCTCTGGCGGAGACGGTGGCCCTTCATCAAGACCATCCGGCT 662
QY 121 CGTGGGCGAGTCTGCTCCCTCGCCACCGGGAGCCGATGGCCCTGAAGGGAGCGCT 180
Db 663 CGTGGGCGAGTCTGCTCCCTCGCCACCGGGAGCCGATGGCCCTGAAGGGAGCGCT 722
QY 181 GCCCGCGCTGCCCCGAGGAGTGGCTTCGAGGGCGGAGGTGAGTACAAACGGGGGCTT 240
Db 723 GCCCGCGCTGCCCCGAGGAGTGGCTTCGAGGGCGGAGGTGAGTACAAACGGGGGCTT 782
QY 241 CCACCTGGCCATCGACGTGA 261
Db 783 CCACCTGGCCATCGACGTGA 803
```

```
RESULT 11
US-10-276-774-950
; Sequence 950, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 950
; LENGTH: 4470
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-276-774-950

Query Match 100.0%; Score 261; DB 13; Length 4470;
Best Local Similarity 100.0%; Pred. No. 2.5e-58;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGGCTCTGCTGGAGGG 60
Db 366 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGGCTCTGCTGGAGGG 425
QY 61 GCTGAGCCTCGGGACGTTCTCTGGCGGAGACGGTGGCCCTTCATCAAGACCATCCGGCT 120
Db 426 GCTGAGCCTCGGGACGTTCTCTGGCGGAGACGGTGGCCCTTCATCAAGACCATCCGGCT 485
QY 121 CGTGGGCGAGTCTGCTCCCTCGCCACCGGGAGCCGATGGCCCTGAAGGGAGCGCT 180
Db 486 CGTGGGCGAGTCTGCTCCCTCGCCACCGGGAGCCGATGGCCCTGAAGGGAGCGCT 545
QY 181 GCCCGCGCTGCCCCGAGGAGTGGCTTCGAGGGCGGAGGTGAGTACAAACGGGGGCTT 240
Db 546 GCCCGCGCTGCCCCGAGGAGTGGCTTCGAGGGCGGAGGTGAGTACAAACGGGGGCTT 605
QY 241 CCACCTGGCCATCGACGTGA 261
Db 606 CCACCTGGCCATCGACGTGA 626
```

```
RESULT 12
US-10-336-603A-25
; Sequence 25, Application US/10336603A
; Publication No. US20040072997A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
```

```
; FILE REFERENCE: 21402-533A
; CURRENT APPLICATION NUMBER: US/10/336,603A
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: 09/746,491
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 10/055,569
; PRIOR FILING DATE: 2001-10-26
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 25
; LENGTH: 4801
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (178)..(3639)
US-10-336-603A-25
```

```
Query Match 100.0%; Score 261; DB 12; Length 4801;
Best Local Similarity 100.0%; Pred. No. 2.5e-58;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGGCTCTGCTGGAGGG 60
Db 543 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGGCTCTGCTGGAGGG 602
QY 61 GCTGAGCCTCGGGACGTTCTCTGGCGGAGACGGTGGCCCTTCATCAAGACCATCCGGCT 120
Db 603 GCTGAGCCTCGGGACGTTCTCTGGCGGAGACGGTGGCCCTTCATCAAGACCATCCGGCT 662
QY 121 CGTGGGCGAGTCTGCTCCCTCGCCACCGGGAGCCGATGGCCCTGAAGGGAGCGCT 180
Db 663 CGTGGGCGAGTCTGCTCCCTCGCCACCGGGAGCCGATGGCCCTGAAGGGAGCGCT 722
QY 181 GCCCGCGCTGCCCCGAGGAGTGGCTTCGAGGGCGGAGGTGAGTACAAACGGGGGCTT 240
Db 723 GCCCGCGCTGCCCCGAGGAGTGGCTTCGAGGGCGGAGGTGAGTACAAACGGGGGCTT 782
QY 241 CCACCTGGCCATCGACGTGA 261
Db 783 CCACCTGGCCATCGACGTGA 803
```

```
RESULT 13
US-10-051-769-4
; Sequence 4, Application US/10051769
; Publication No. US20030044811A1
; GENERAL INFORMATION:
; APPLICANT: McKINNON, Randy D.
; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION
; FILE REFERENCE: 268/260 (RWJ-00-37)
; CURRENT APPLICATION NUMBER: US/10/051,769
; CURRENT FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,160
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 263
; TYPE: DNA
; ORGANISM: Sprague Dawley rat
US-10-051-769-4
```

```
Query Match 79.8%; Score 208.2; DB 15; Length 263;
Best Local Similarity 87.4%; Pred. No. 1.4e-44;
Matches 228; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGGCTCTGCTGGAGGG 60
Db 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCTTTTTTTTTTGGAGGG 60
QY 61 GCTGAGCCTCGGGACGTTCTCTGGCGGAGACGGTGGCCCTTCATCAAGACCATCCGGCT 120
```

Db 61 GCTGAGCTCGCGAGCTGTTCCTGGGTGACACCGCTGCCTACATCAAGACCATCCGGCT 120
QY 121 CTTGGCGCCAGTCTCCCTCGCCACCGGAGCCGATGGCCCTGAAGGGAGCGCT 180
Db 121 GGTGGCGCCGCTGTGGCTTCGGGACCGCGAGCCCGACGAACCCGATGGGAGCTCT 180
QY 181 GCCCGCGCTGCTCCCGAGGAGTGGCTTCGAGGCGGAGGTGAGTACAACCGGGGCTT 240
Db 181 GCCCGCACCTGCGCGGGGAGCTGSCCTTTGAGGCGGAGTGGAGTACAACCGCGGCTT 240
QY 241 CCACCTGGCCATCGACGTGA 261
Db 241 CCACCTGGCCATCGACGTGA 261

RESULT 14

US-10-224-624-4
; Sequence 4, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: McKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GLITEN
; FILE REFERENCE: 54704.8059.US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; PRIOR FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 263
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-224-624-4

Query Match 79.8%; Score 208.2; DB 15; Length 263;
Best Local Similarity 87.4%; Pred. No. 1.4e-44;
Matches 228; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 1 GATCAAGTGGAGTTCGAGGAGCTCTGCGAGACCAAGACGGCGCGGCGCTCTGAGGG 60
Db 1 GATCAAGTGGAGTTCGAGGAGCTCTGCGAGACCAAGACGGCGCTTTTTTTTGGAGGG 60
QY 61 GCTGAGCTCGGAGCTGTTCCTGGCGAGACGCTGCTTCATCAAGACCATCCGGCT 120
Db 61 GCTGAGCTCGGAGCTGTTCCTGGGTGACACCGTGCCTACATCAAGACCATCCGGCT 120
QY 121 CGTGGGCGAGTGTGCTCGCCACCGCGAGCCCGATGGCCCTGAAGGGAGCGCT 180
Db 121 GGTGGGCGCGTGTGTGGCTTCGGGCACCGCGAGCCCGACGAACCCGATGGGAGCTCT 180
QY 181 GCCCGCGCTGCTCCCGAGGAGCTGGCTTCGAGGCGGAGGTGAGTACAACCGGGGCTT 240
Db 181 GCCCGCACCTGCGCGGGGAGCTGSCCTTTGAGGCGGAGTGGAGTACAACCGCGGCTT 240
QY 241 CCACCTGGCCATCGACGTGA 261
Db 241 CCACCTGGCCATCGACGTGA 261

RESULT 15

US-10-156-761-3629
; Sequence 3629, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3629
; LENGTH: 11058
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(11058)
US-10-156-761-3629
Query Match 19.6%; Score 51.2; DB 15; Length 11058;
Best Local Similarity 51.3%; Pred. No. 0.00029;
Matches 119; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
QY 26 CTGCAGACCAAGACGGCGCGCTGCTGGAGGGCTGAGCTCGGGAGCGTGTTCCTG 85
Db 2830 CTGGCTCCACCTGGCGCCCGCTGCCGAGTACATGGTCCGGCCCGCTTCGTCGC 2889
QY 86 GCGGAGCGTGCCTTCATCAAGACCATCCGCTCGTGGCGCCAGTCTGCTCGGCC 145
Db 2890 CTCGACGGCTGCCCTTGACCAACGCGAAGCTCGACAAGCGCCCTGCCCGCCC 2949
QY 146 ACCGGGAGCGGATGGCTGAGGGAGGGCTGCGCGCGCTGCCCGCTGCCCGAGGCTG 205
Db 2950 GCGCAGGACGCGCTCGCGCGCCGCGCCACGTCGCGCCCGCACCGTCCCGAGGCGG 3009
QY 206 GCCTTCGAGGGAGGTGAGTACAAACGGGGCTTCCACCTGGCCATCGACG 257
Db 3010 ATCGCGAGGTGCGCGCGACGCTCTCGCCCTCGACCGGCTCGGTCTGAGG 3061
Search completed: July 18, 2004, 14:47:31
Job time : 463.82 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 11:31:29 ; Search time 435.57 Seconds

(without alignments)
2545:580 Million cell updates/sec

Title: US-10-051-769-2

Perfect score: 261

Sequence: 1 gatcaagggtgagttcgagg.....cactggccatcgactgga 261

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004as:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	261	100.0	261	7	ABX95679 Human Gli
2	261	100.0	261	9	AAD61560 Human Gli
3	261	100.0	873	7	ABX95680 Human Gli
4	261	100.0	960	7	ABX95678 Human gen
5	261	100.0	960	9	AAD61559 Human Gli
6	261	100.0	1092	9	AD607789 Novel cod
7	261	100.0	2639	9	AD862530 Human cDN
8	261	100.0	3832	9	AAD61564 Human Gli
9	261	100.0	4450	4	AAD61568 Human pol
10	261	100.0	4470	4	ABA09174 Human sec
11	261	100.0	4470	4	AAX52592 Human pol
12	261	100.0	4470	9	AD609891 Novel DNA
13	261	100.0	4801	9	ADDA9052 Human NOV
14	208.2	79.8	263	7	ABX95681 Rat GliTE
15	208.2	79.8	263	9	AAD61561 Human Gli
16	169.8	65.1	1732	6	ABQ16934 Oligonuel
17	169.8	65.1	1732	6	ABQ16935 Oligonuel
18	137	52.5	1732	6	ABQ16937 Oligonuel
19	137	52.5	1732	6	ABQ16936 Oligonuel
20	51.6	19.8	6436	2	AAX86965 DNA confa
21	48.6	18.6	1476	6	ABQ91125 M. capsul
22	47.8	18.3	2382	9	ADC30095 Human nov
23	47.8	18.3	2710	6	ABN85324 Human cyt

24	47	18.0	3847	6	ABK87764 Human cDN
25	47	18.0	9994	4	AAC85191 S. avermi
26	47	18.0	10232	7	ACA37617 Prokaryot
27	46.6	17.9	2000	7	ADA71938 Rice gene
28	46	17.6	109519	5	AAS08693 Micromono
29	45.2	17.3	2047	5	AAS83393 DNA encod
30	45.2	17.3	2499	5	AAS85836 DNA encod
31	45.2	17.3	4263	5	AAS93470 DNA encod
32	45.2	17.3	4440	3	AAA92169 Human POL
33	45.2	17.3	4465	6	AAL38925 Human pol
34	45.2	17.3	4467	3	AAC76835 Human ORF
35	45.2	17.3	4786	5	AAS66446 DNA encod
36	45.2	17.3	75384	4	AAS85590 Human imm
37	44.8	17.2	2107	3	AAC86412 Wheat sta
38	44.8	17.2	2107	3	AAC86434 Wheat sta
39	44.4	17.0	2622	7	ACA26991 Prokaryot
40	44.4	17.0	23673	6	ABZ75344 Human R11
41	44	16.9	1164	7	ACA38173 Prokaryot
42	43.8	16.8	1008	8	ADA48337 Rice gene
43	43.8	16.8	3959	6	ABI99206 Mouse isc
44	43.8	16.8	4347	7	ACA42414 Prokaryot
45	43.8	16.8	5811	6	ABS78655 S. mactrom

ALIGNMENTS

RESULT 1

ABX95679

ID ABX95679 standard; cDNA; 261 BP.

XX

AC ABX95679;

XX

DT 30-JUN-2003 (first entry)

XX

DE Human GliTEN expressed sequence tag.

XX

Human; ss; EST; chromosome 10q25; glioblastoma multiforme; brain cancer;

KW GliTEN; GMB; expressed sequence tag.

XX

OS Homo sapiens.

XX

PN US2003044811-A1.

XX

PD 06-MAR-2003.

XX

PF 20-OCT-2001; 2001US-00051769.

XX

PR 20-OCT-2000; 2000US-0242160P.

XX

PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

XX

PI McKinnon RD;

XX

DR WPI; 2003-418995/39.

XX

PT New isolated nucleic acid representing a gene product associated with Glioblastoma Multiforme, designated as GliTEN, useful as a molecular marker for tumor cell identification and classification.

PT

XX

PS Claim 1; Page 7; 11pp; English.

XX

CC The invention relates to an isolated nucleic acid representing a gene product associated with Glioblastoma Multiforme, designated as GliTEN, comprising the human EST (expressed sequence tag) appearing as ABX95678, or a sequence that hybridises under stringent conditions to the EST, or its complement. Also included are a probe for use in identifying a patient at risk for progression into the malignant phenotype, comprising the nucleic acids detailed above, detecting whether a patient is at risk for progression into Glioblastoma Multiforme (comprising: (a) providing a sample from a patient; (b) adding the probe to the sample or performing a PCR analysis using the primers appearing as ABX95682 and ABX95683; (c) analysing levels of mRNA bound with the probe; and (d) treating a control

CC sample to assess the level of mRNA in a control sample, where the
 CC presence of increased levels of mRNA expression in the sample in an
 CC amount higher than the control sample indicates risk for progression into
 CC Glioblastoma Multiforme), and a kit for use in detecting whether a
 CC patient is at risk for progression into Glioblastoma Multiforme
 CC (comprising the probe or primers). The probe or the kit is useful for
 CC detecting whether a patient is at risk for progression into Glioblastoma
 CC Multiforme (a type of brain cancer). The EST useful as molecular marker
 CC for tumour cell identification and classification and for diagnosing or
 CC identifying candidates at risk for progression into a malignant phenotype
 CC especially in brain cancer therapy, treatment and diagnosis. The present
 CC sequence is the human GliTEN expressed sequence tag
 XX
 SQ Sequence 261 BP; 40 A; 81 C; 102 G; 38 T; 0 U; 0 Other;

Query Match 100.0%; Score 261; DB 7; Length 261;
 Best Local Similarity 100.0%; Pred. No. 1.6e-40;
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GATCAAGGTGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCCGGCGCCCTGCTGGAGGG 60
 DB 1 GATCAAGGTGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCCGGCGCCCTGCTGGAGGG 60
 QY 61 GCTGAGCCTGCGGAGCGTTCCTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
 DB 61 GCTGAGCCTGCGGAGCGTTCCTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
 QY 121 CGTGGGCGCAGTTCGTGCCCTCGGCCACCGGGGAGCCCGATGCCCTGAAGGGGAGCGCT 180
 DB 121 CGTGGGCGCAGTTCGTGCCCTCGGCCACCGGGGAGCCCGATGCCCTGAAGGGGAGCGCT 180
 QY 181 GCCCGCCCGCTGCCCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT 240
 DB 181 GCCCGCCCGCTGCCCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT 240
 QY 241 CCACCTGGCCATCGACGTGGA 261
 DB 241 CCACCTGGCCATCGACGTGGA 261

RESULT 2
 AAD61560
 ID AAD61560 standard; cDNA; 261 BP.
 AC AAD61560;
 XX
 DT 15-JAN-2004 (first entry)
 XX
 DE Human GliTEN EST clone cDNA #1.
 XX
 KW Human; glioblastoma multiforme; GBM; GliTEN; brain cancer; diagnosis;
 KW gene therapy; expressed sequence tag; ss.
 XX
 OS Homo sapiens.
 XX
 PN US2003108915-A1.
 XX
 PD 12-JUN-2003.
 XX
 PF 20-AUG-2002; 2002US-00224624.
 XX
 PR 20-OCT-2000; 2000US-0242160P.
 PR 20-OCT-2001; 2001US-00051769.
 XX
 PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
 XX
 PI McKinnon RD;
 XX
 DR WPI; 2003-810848/76.
 XX
 PT Novel glioblastoma multiforme associated protein GliTEN useful for
 PT treating glioblastoma multiforme and diagnosing brain cancer.
 XX

PS Claim 38; Page 11; Opp; English.
 XX
 CC The present invention provides novel glioblastoma multiforme (GBM)
 CC associated protein GliTEN useful for treating glioblastoma multiforme and
 CC diagnosing brain cancer. The invention is useful in brain cancer therapy,
 CC treatment and diagnosis. The invention is also useful in gene therapy.
 CC The present sequence is human GliTEN EST clone cDNA. This EST clone is
 CC used in the exemplification of the invention
 XX
 SQ Sequence 261 BP; 40 A; 81 C; 102 G; 38 T; 0 U; 0 Other;
 Query Match 100.0%; Score 261; DB 9; Length 261;
 Best Local Similarity 100.0%; Pred. No. 1.6e-40;
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GATCAAGGTGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCCGGCGCCCTGCTGGAGGG 60
 DB 1 GATCAAGGTGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCCGGCGCCCTGCTGGAGGG 60
 QY 61 GCTGAGCCTGCGGAGCGTTCCTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
 DB 61 GCTGAGCCTGCGGAGCGTTCCTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
 QY 121 CGTGGGCGCAGTTCGTGCCCTCGGCCACCGGGGAGCCCGATGCCCTGAAGGGGAGCGCT 180
 DB 121 CGTGGGCGCAGTTCGTGCCCTCGGCCACCGGGGAGCCCGATGCCCTGAAGGGGAGCGCT 180
 QY 181 GCCCGCCCGCTGCCCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT 240
 DB 181 GCCCGCCCGCTGCCCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT 240
 QY 241 CCACCTGGCCATCGACGTGGA 261
 DB 241 CCACCTGGCCATCGACGTGGA 261
 RESULT 3
 ABX95680
 ID ABX95680 standard; DNA; 873 BP.
 XX
 AC ABX95680;
 XX
 DT 30-JUN-2003 (first entry)
 XX
 DE Human GliTEN open reading frame.
 XX
 KW Human; ds; chromosome 10q25; glioblastoma multiforme; brain cancer;
 KW GliTEN; GMB.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..873
 FT /*tag= a
 FT /partial
 FT /product= "GliTEN"
 FT /note= "No stop codon shown"
 XX
 PN US2003044811-A1.
 XX
 PD 06-MAR-2003.
 XX
 PF 20-OCT-2001; 2001US-00051769.
 XX
 PR 20-OCT-2000; 2000US-0242160P.
 XX
 PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
 XX
 PI McKinnon RD;
 XX
 DR WPI; 2003-418995/39.
 XX
 PT New isolated nucleic acid representing a gene product associated with

PT Glioblastoma Multiforme, designated as GliTEN, useful as a molecular
PT marker for tumor cell identification and classification.
XX
PS Disclosure; Page 7; 11pp; English.
XX
XX The invention relates to an isolated nucleic acid representing a gene
CC product associated with Glioblastoma Multiforme, designated as GliTEN,
CC comprising the human EST (expressed sequence tag) appearing as ABX95678,
CC or a sequence that hybridises under stringent conditions to the EST, or
CC its complement. Also included are a probe for use in identifying a
CC patient at risk for progression into the malignant phenotype, comprising
CC the nucleic acids detailed above, detecting whether a patient is at risk
CC for progression into Glioblastoma Multiforme (comprising: (a) providing a
CC sample from a patient; (b) adding the probe to the sample or performing
CC PCR analysis using the primers appearing as ABX95682 and ABX95683; (c)
CC analysing levels of mRNA bound with the probe; and (d) treating a control
CC sample to assess the level of mRNA in a control sample, where the
CC presence of increased levels of mRNA expression in the sample in an
CC amount higher than the control sample indicates risk for progression into
CC Glioblastoma Multiforme), and a kit for use in detecting whether a
CC patient is at risk for progression into Glioblastoma Multiforme
CC (comprising the probe or primers). The probe or the kit is useful for
CC detecting whether a patient is at risk for progression into Glioblastoma
CC Multiforme (a type of brain cancer). The EST useful as molecular marker
CC for tumour cell identification and classification and for diagnosing or
CC identifying candidates at risk for progression into a malignant phenotype
CC especially in brain cancer therapy, treatment and diagnosis. The present
CC sequence is the open reading frame from the GliTEN gene
XX
SQ Sequence 873 BP; 137 A; 304 C; 279 G; 153 T; 0 U; 0 Other;

Query Match 100.0%; Score 261; DB 7; Length 873;
Best Local Similarity 100.0%; Pred. No. 1.5e-40;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 60
DB 366 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 425

QY 61 GCTGAGCCTCGCGGAGCTGTTCTCTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
DB 426 GCTGAGCCTCGCGGAGCTGTTCTCTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 485

QY 121 CGTGGCGCAGTCTGCTCCCTCGCCACCGGGAGCCCGATGGCCCTGAAGGGAGCGCT 180
DB 486 CGTGGCGCAGTCTGCTCCCTCGCCACCGGGAGCCCGATGGCCCTGAAGGGAGCGCT 545

QY 181 GCCCGCGCTGCTCCCGAGGAGCTGGCTTCGAGGCGGAGTGGAGTACAAACGGGGGCTT 240
DB 546 GCCCGCGCTGCTCCCGAGGAGCTGGCTTCGAGGCGGAGTGGAGTACAAACGGGGGCTT 605

QY 241 CCACCTGGCCATCGACGTGA 261
DB 606 CCACCTGGCCATCGACGTGA 626

RESULT 4
ABX95678
ID ABX95678 standard; DNA; 960 BP.
XX
AC ABX95678;
XX
XX 30-JUN-2003 (first entry)
DE Human gene encoding GliTEN.
XX
XX Human; ds; gene; chromosome 10q25; glioblastoma multiforme; brain cancer;
KW GliTEN; GMB.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH 30..906
FT CDS

FT
FT
XX
PN
XX
XX
PD
XX
XX
PF 20-OCT-2001; 2001US-00051769.
XX
XX 20-OCT-2000; 2000US-0242160P.
XX
XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX
XX McKinnon RD;
XX
XX WPI; 2003-418995/39.
XX
XX New isolated nucleic acid representing a gene product associated with
PT Glioblastoma Multiforme, designated as GliTEN, useful as a molecular
PT marker for tumor cell identification and classification.
XX
XX Example 2; Fig 1; 11pp; English.
XX
XX The invention relates to an isolated nucleic acid representing a gene
CC product associated with Glioblastoma Multiforme, designated as GliTEN,
CC comprising the human EST (expressed sequence tag) appearing as ABX95678,
CC or a sequence that hybridises under stringent conditions to the EST, or
CC its complement. Also included are a probe for use in identifying a
CC patient at risk for progression into the malignant phenotype, comprising
CC the nucleic acids detailed above, detecting whether a patient is at risk
CC for progression into Glioblastoma Multiforme (comprising: (a) providing a
CC sample from a patient; (b) adding the probe to the sample or performing
CC PCR analysis using the primers appearing as ABX95682 and ABX95683; (c)
CC analysing levels of mRNA bound with the probe; and (d) treating a control
CC sample to assess the level of mRNA in a control sample, where the
CC presence of increased levels of mRNA expression in the sample in an
CC amount higher than the control sample indicates risk for progression into
CC Glioblastoma Multiforme), and a kit for use in detecting whether a
CC patient is at risk for progression into Glioblastoma Multiforme
CC (comprising the probe or primers). The probe or the kit is useful for
CC detecting whether a patient is at risk for progression into Glioblastoma
CC Multiforme (a type of brain cancer). The EST useful as molecular marker
CC for tumour cell identification and classification and for diagnosing or
CC identifying candidates at risk for progression into a malignant phenotype
CC especially in brain cancer therapy, treatment and diagnosis. The present
CC sequence is a region of human chromosome 10 (q25) containing the GliTEN
CC gene
XX
SQ Sequence 960 BP; 146 A; 324 C; 329 G; 161 T; 0 U; 0 Other;

Query Match 100.0%; Score 261; DB 7; Length 960;
Best Local Similarity 100.0%; Pred. No. 1.5e-40;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 60
DB 396 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 455

QY 61 GCTGAGCCTCGCGGAGCTGTTCTCTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
DB 456 GCTGAGCCTCGCGGAGCTGTTCTCTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 515

QY 121 CGTGGCGCAGTCTGCTCCCTCGCCACCGGGAGCCCGATGGCCCTGAAGGGAGCGCT 180
DB 516 CGTGGCGCAGTCTGCTCCCTCGCCACCGGGAGCCCGATGGCCCTGAAGGGAGCGCT 575

QY 181 GCCCGCGCTGCTCCCGAGGAGCTGGCTTCGAGGCGGAGTGGAGTACAAACGGGGGCTT 240
DB 576 GCCCGCGCTGCTCCCGAGGAGCTGGCTTCGAGGCGGAGTGGAGTACAAACGGGGGCTT 635

QY 241 CCACCTGGCCATCGACGTGA 261
DB 636 CCACCTGGCCATCGACGTGA 656

RESULT 5

AAD61559
ID AAD61559 standard; cDNA; 960 BP.
XX
AC AAD61559;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human GliTEN partial cDNA.
XX
KW Human; glioblastoma multiforme; GBM; GliTEN; brain cancer; diagnosis;
KW gene therapy; chromosome 10; gene; ss.
XX
OS Homo sapiens.
XX
PN US2003108915-A1.
XX
PD 12-JUN-2003.
XX
PF 20-AUG-2002; 2002US-00224624.
XX
PR 20-OCT-2000; 2000US-0242160P.
XX
PR 20-OCT-2001; 2001US-00051769.
XX
PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX
PI McKinnon RD;
XX
DR WPI; 2003-810848/76.
XX
PT Novel glioblastoma multiforme associated protein GliTEN useful for
PT treating glioblastoma multiforme and diagnosing brain cancer.
XX
PS Example 2; Fig 1; Opp; English.
XX
CC The present invention provides novel glioblastoma multiforme (GBM)
CC associated protein GliTEN useful for treating glioblastoma multiforme and
CC diagnosing brain cancer. The invention is useful in brain cancer therapy,
CC treatment and diagnosis. The invention is also useful in gene therapy.
CC The present sequence is human GliTEN partial cDNA. The human GliTEN gene
CC is located on chromosome 10. This partial cDNA is used in the
CC exemplification of the invention
XX
SQ Sequence 960 BP; 146 A; 324 C; 329 G; 161 T; 0 U; 0 Other;
Query Match 100.0%; Score 261; DB 9; Length 960;
Best Local Similarity 100.0%; Pred. No. 1.5e-40;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCACACCAAGACGCGCGCGCCCTGCTGGAGGG 60
DB 396 GATCAAGTGGAGTTCGAGGAGCTGCTGCACACCAAGACGCGCGCGCCCTGCTGGAGGG 455
QY 61 GCTGAGCTCGGACGCTGTTCTCTGGGCGAGACGCTGCCCTTCATCAAGACCATCCGCGCT 120
DB 456 GCTGAGCTCGGACGCTGTTCTCTGGGCGAGACGCTGCCCTTCATCAAGACCATCCGCGCT 515
QY 121 CGTGGGCGCAGTCGTGCCCTCGGCCACCGGGAGCCCGATGCGCCCTGAAGGGGAGCGCT 180
DB 516 CGTGGGCGCAGTCGTGCCCTCGGCCACCGGGAGCCCGATGCGCCCTGAAGGGGAGCGCT 575
QY 181 GCCCGCGCTGCCCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGAGTACAAAGGGGGCTT 240
DB 576 GCCCGCGCTGCCCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGAGTACAAAGGGGGCTT 635
QY 241 CCACCTGCCATCGACGTGGA 261
DB 636 CCACCTGCCATCGACGTGGA 656

RESULT 6

ADE07789
ID ADE07789 standard; DNA; 1092 BP.
XX
AC ADE07789;
XX
DT 29-JAN-2004 (first entry)
XX
DE Novel coding sequence (useful for identifying genetic disorders) #855.
XX
KW novel gene; novel protein; tissue marker; molecular weight marker;
KW chromosome marker; genetic disorder; gene; ds.
XX
OS Unidentified.
XX
PN WO20003054152-A2.
XX
PD 03-JUL-2003.
XX
PF 10-DEC-2002; 2002WO-US039555.
XX
PR 10-DEC-2001; 2001US-0339739P.
PR 11-DEC-2001; 2001US-0339453P.
PR 14-MAR-2002; 2002US-0365091P.
PR 14-MAR-2002; 2002US-0365384P.
PR 12-APR-2002; 2002US-0372381P.
PR 12-APR-2002; 2002US-0372615P.
PR 22-APR-2002; 2002US-00128558.
PR 24-APR-2002; 2002US-0376045P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;
PI Ma Y, Wang D, Chen R, Xu C, Boyle BU;
XX
DR WPI; 2003-569235/53.
DR P-PSDB; ADE08700.
XX
PT New polynucleotides, useful for expressing recombinant proteins for
PT analysis, characterization or therapeutic use, or as markers for tissues
PT in which the corresponding protein is preferentially expressed.
XX
PS Claim 1; SEQ ID NO 855; 1177pp; English.
XX
CC The invention comprises the amino acid and coding sequences of novel
CC proteins. The DNA and protein sequences of the invention are useful as:
CC markers for tissues in which the corresponding protein is preferentially
CC expressed; as molecular weight markers on gels; as chromosome markers or
CC tags; to identify chromosomes or to map related gene positions; and to
CC compare with endogenous DNA sequences in patients to identify potential
CC genetic disorders. The present DNA sequence represents a gene of the
CC invention.
XX
SQ Sequence 1092 BP; 190 A; 377 C; 322 G; 203 T; 0 U; 0 Other;
Query Match 100.0%; Score 261; DB 9; Length 1092;
Best Local Similarity 100.0%; Pred. No. 1.5e-40;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCACACCAAGACGCGCGCGCCCTGCTGGAGGG 60
DB 366 GATCAAGTGGAGTTCGAGGAGCTGCTGCACACCAAGACGCGCGCGCCCTGCTGGAGGG 425
QY 61 GCTGAGCTCGGACGCTGTTCTCTGGGCGAGACGCTGCCCTTCATCAAGACCATCCGCGCT 120
DB 426 GCTGAGCTCGGACGCTGTTCTCTGGGCGAGACGCTGCCCTTCATCAAGACCATCCGCGCT 485
QY 121 CGTGGGCGCAGTCGTGCCCTCGGCCACCGGGAGCCCGATGCGCCCTGAAGGGGAGCGCT 180
DB 486 CGTGGGCGCAGTCGTGCCCTCGGCCACCGGGAGCCCGATGCGCCCTGAAGGGGAGCGCT 545
QY 181 GCCCGCGCTGCCCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGAGTACAAAGGGGGCTT 240

```
Db 546 GCCCGCGCTGCCCGAGGAGTGGCTTCGAGGCGGAGGTGGAGTACACAGGGGGCTT 605
QY 241 CCACCTGGCCATCGAGTGA 261
    |||||
Db 606 CCACCTGGCCATCGAGTGA 626
    |||||

RESULT 7
ADB62530
ID ADB62530 standard; cDNA; 2639 BP.
XX
AC ADB62530;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human cDNA encoding clone HHDPC20088160.
XX
KW Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW tissue regeneration; cell regeneration; membrane protein;
KW signal transduction-related protein; transcription-related protein;
KW osteoporosis; neurological disease; cancer; tumour.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 214..2639
FT /tag= a
FT /partial
FT /product= "Clone HHDPC20088160 protein"
XX
PN EP1308459-A2.
XX
PD 07-MAY-2003.
XX
PF 28-MAR-2002; 2002EP-00007401.
XX
PR 05-NOV-2001; 2001JP-00379298.
PR 25-JAN-2002; 2002US-00350978.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-450961/43.
DR P-PSDB; ADB64500.
XX
FT New polynucleotides and polypeptides, useful for developing a diagnostic
FT marker or medicines for regulation of their expression and activity, or
FT as targets of gene therapy.
XX
PS Claim 1; Page; 222pp; English.
XX
CC The invention discloses a polynucleotide comprising a sequence selected
CC from 1970 fully defined nucleotide sequences which encode novel
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC or its partial peptide, an antibody binding to the polypeptide or peptide
CC of the polynucleotide, immunologically assaying the polypeptide or
CC peptide of the polynucleotide by contacting the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesising the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotides and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease-related proteins and genes
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
```

```
CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a cDNA of the invention. Note: Some of the sequence
CC data for this patent is not represented in the printed specification, but
CC is based on sequence information supplied by the European Patent Office.
XX
SQ Sequence 2639 BP; 657 A; 702 C; 719 G; 561 T; 0 U; 0 Other;

Query Match 100.0%; Score 261; DB 9; Length 2639;
Best Local Similarity 100.0%; Pred. No. 1.4e-40;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGGTGGAGTTCGAGGAGTCTGCAGACCAAGACGCGCGGCGCTGCTGGAGGG 60
    |||||
Db 579 GATCAAGGTGGAGTTCGAGGAGTCTGCAGACCAAGACGCGCGGCGCTGCTGGAGGG 638
    |||||
QY 61 GCTGAGCCTGCGGACGCTGCTTCTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
    |||||
Db 639 GCTGAGCCTGCGGACGCTGCTTCTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 698
    |||||
QY 121 CBTGCGGCCAGTCTGTGCCCTCGGCCACCGGGAGCCCGATGCGCCCTGAAGGGGAGCGCT 180
    |||||
Db 699 CBTGCGGCCAGTCTGTGCCCTCGGCCACCGGGAGCCCGATGCGCCCTGAAGGGGAGCGCT 758
    |||||
QY 181 GCCCGCGCTGCCCGAGGAGTGGCCTTCGAGGCGGAGTGGAGTACACAGGGGGCTT 240
    |||||
Db 759 GCCCGCGCTGCCCGAGGAGTGGCCTTCGAGGCGGAGTGGAGTACACAGGGGGCTT 818
    |||||
QY 241 CCACCTGGCCATCGAGTGA 261
    |||||
Db 819 CCACCTGGCCATCGAGTGA 839
    |||||

RESULT 8
AAD61564
ID AAD61564 standard; cDNA; 3832 BP.
XX
AC AAD61564;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human GliTEN cDNA.
XX
KW Human; glioblastoma multiforme; GBM; GliTEN; brain cancer; diagnosis;
KW gene therapy; chromosome 10; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 178..3642
FT /tag= a
FT /product= "Human GliTEN protein"
FT /note= "The CDS is specifically claimed in claim 1"
FT primer_bind 1011..1032
FT /tag= b
FT /bound_moiety= "5' primer #843"
FT primer_bind 1405..1429
FT /tag= c
FT /bound_moiety= "5' primer #1405"
FT primer_bind 2083..2108
FT /tag= d
FT /bound_moiety= "5' primer #2083"
FT primer_bind 2356..2378
FT /tag= e
FT /bound_moiety= "5' primer #2355"
FT primer_bind 3625..3650
FT /tag= f
FT /bound_moiety= "3' primer #3636"
XX
PN US2003108915-A1.
XX
PD 12-JUN-2003.
```


DE Human secreted protein homologue-encoding cDNA, SEQ ID NO:950.
XX
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytosstatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnerary; antiulcer; ss.
XX
OS Homo sapiens.
XX
XX WO2001571188-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US003800.
XX
XX 03-FEB-2000; 2000US-00496914.
XX
XX 27-APR-2000; 2000US-00560875.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-457740/49.
XX
XX P-PSDB; ABB11930.
XX
XX Human proteins and DNA encoding sequences useful for preventing, treating
XX or ameliorating a medical condition in a mammalian subject e.g. arthritis
XX and cancer.
XX
XX Claim 1; Page 813-814; 1963pp; English.
XX
XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
XX sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
XX invention also relates to vectors and recombinant host cells comprising a
XX nucleotide of the invention, methods of producing the novel polypeptides,
XX antibodies against the polypeptides, methods of detecting the nucleotides
XX or polypeptides in a sample, and methods of identifying compounds which
XX bind to polypeptides of the invention. Although novel, many of the
XX polypeptides of the invention have homology to known proteins, thereby
XX giving an insight into their probable biological activities, and hence
XX potential therapeutic applications. The polypeptides of the invention may
XX have various activities, including cytokine, cell proliferation or cell
XX differentiation activities; stem cell growth factor activity;
XX haematopoiesis regulatory activity; tissue growth activity;
XX immunomodulatory activity; activin- or inhibin-related activities;
XX chemotactic or chemokinetic activities; haemostatic, thrombotic or
XX thrombolytic activities; receptor or ligand activities; or may be
XX involved in oncogenesis, cancer cell proliferation or metastasis.
XX Depending on their biological activities, polypeptides and nucleotides of
XX the invention are useful for preventing, treating or ameliorating medical
XX conditions, e.g., by protein or gene therapy. Such conditions include
XX cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
XX disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
XX proliferative retinopathy, atherosclerosis, coronary heart disease,
XX arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
XX vascular growth. Polypeptides involved with tissue regeneration and
XX repair (or nucleic acids encoding them) may be used to promote wound
XX healing (e.g., of burns, incisions and ulcers), while those with
XX immunomodulatory activities may be used in the treatment of viral,
XX bacterial and fungal infections in addition to immune disorders.
XX Polypeptides with growth factor activity may be used in cell cultures to
XX promote cell growth. For example, such polypeptides may be used to
XX manipulate stem cells in culture to give rise to neuroepithelial cells
XX that can be used to augment or replace cells damaged by illness,

CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a cDNA encoding a
CC novel human polypeptide of the invention
XX
SQ Sequence 4470 BP; 1254 A; 991 C; 1054 G; 1171 T; 0 U; 0 Other;
Query Match 100.0%; Score 261; DB 4; Length 4470;
Best Local Similarity 100.0%; Pred. No. 1.3e-40;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCGCTGCTGGAGGG 60
Db 366 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCGCTGCTGGAGGG 425
QY 61 GCTGAGCCTCGGGACGTTCTCTGGGCGAGACGCTTCAATCAAGACCATCCGGCT 120
Db 426 GCTGAGCCTCGGGACGTTCTCTGGGCGAGACGCTTCAATCAAGACCATCCGGCT 485
QY 121 CGTGGCGCAGTGTGCTCCCTCGGCCACCGGGGAGCCCGTGAAGGGGAGCGCT 180
Db 486 CGTGGCGCAGTGTGCTCCCTCGGCCACCGGGGAGCCCGTGAAGGGGAGCGCT 545
QY 181 GCCCGCGCTCGCCCGGAGGAGCTGGCTTCGAGGCGGAGTGAAGTACACGCGGGCTT 240
Db 546 GCCCGCGCTCGCCCGGAGGAGCTGGCTTCGAGGCGGAGTGAAGTACACGCGGGCTT 605
QY 241 CCACCTGGCCATCGACGTGGA 261
Db 606 CCACCTGGCCATCGACGTGGA 626
RESULT 11
AAK52592
ID AAK52592 standard; cDNA; 4470 BP.
AC AAK52592;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 2121.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
XX Homo sapiens.
XX
XX WO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US004098.
XX
XX 03-FEB-2000; 2000US-00496914.
XX
XX 27-APR-2000; 2000US-00560875.
XX
XX 20-JUN-2000; 2000US-00598075.
XX
XX 19-JUL-2000; 2000US-00620325.
XX
XX 01-SEP-2000; 2000US-00654936.
XX
XX 15-SEP-2000; 2000US-00663561.
XX
XX 20-OCT-2000; 2000US-00693325.
XX
XX 30-NOV-2000; 2000US-00728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
XX Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
XX
XX P-PSDB; AAM79459.
XX

PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.

XX Claim 1; Page 4496-4497; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAW78323-AAW80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
XX (AAK52582) and 3666 (AAW80020) are omitted as the relevant pages from the
XX sequence listing were missing at the time of publication

SQ Sequence 4470 BP; 1254 A; 991 C; 1054 G; 1171 T; 0 U; 0 Other;

Query Match 100.0%; Score 261; DB 4; Length 4470;
Best Local Similarity 100.0%; Pred. No. 1.3e-40;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGGCGCTGCTGGAGG 60
DB 366 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGGCGCTGCTGGAGG 425

QY 61 GCTGAGCCTCGCGGACGTGTTCTCTGGCGAGACGGTGCCTTTCATCAAGACCATCCGGCT 120
DB 426 GCTGAGCCTCGCGGACGTGTTCTCTGGCGAGACGGTGCCTTTCATCAAGACCATCCGGCT 485

QY 121 CQTGCGGCCAGTCTGTCCTCGGCCACCGGGAGCCGATGCCCTGAAGGGGAGCGCT 180
DB 486 CQTGCGGCCAGTCTGTCCTCGGCCACCGGGAGCCGATGCCCTGAAGGGGAGCGCT 545

QY 181 GCCCGCCGCTGCCCGGAGGAGCTGCTTCAGGCGGAGGTGGAGTACAACGGGGGCTT 240
DB 546 GCCCGCCGCTGCCCGGAGGAGCTGCTTCAGGCGGAGGTGGAGTACAACGGGGGCTT 605

QY 241 CCACCTGGCCATCGACGTGGA 261
DB 606 CCACCTGGCCATCGACGTGGA 626

RESULT 12
ADE09891
ID ADE09891 standard; DNA; 4470 BP.

XX AC ADE09891;

XX DT 29-JAN-2004 (first entry)

XX DE Novel DNA-related contig nucleotide sequence #613.

XX KW novel gene; novel protein; tissue marker; molecular weight marker;

XX KW chromosome marker; genetic disorder; contig; ds.

XX OS Unidentified.

XX PN WO2003054152-A2.

XX PD 03-JUL-2003.

XX PF 10-DEC-2002; 2002WO-US039555.

XX PR 10-DEC-2001; 2001US-0339739P.

XX PR 11-DEC-2001; 2001US-0339453P.

XX PR 14-MAR-2002; 2002US-0365091P.

XX PR 14-MAR-2002; 2002US-0365384P.

XX PR 12-APR-2002; 2002US-0372381P.

XX PR 12-APR-2002; 2002US-0372615P.

PR 22-APR-2002; 2002US-00128558.
PR 24-APR-2002; 2002US-0376045P.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
XX PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;
XX PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;
XX WPI; 2003-569235/53.

XX New polynucleotides, useful for expressing recombinant proteins for
XX analysis, characterization or therapeutic use, or as markers for tissues
XX in which the corresponding protein is preferentially expressed.

PS Disclosure; SEQ ID NO 2435; 1177pp; English.

XX The invention comprises the amino acid and coding sequences of novel
XX proteins. The DNA and protein sequences of the invention are useful as:
XX markers for tissues in which the corresponding protein is preferentially
XX expressed; as molecular weight markers on gels; as chromosome markers or
XX tags; to identify chromosomes or to map related gene positions; and to
XX compare with endogenous DNA sequences in patients to identify potential
XX genetic disorders. The present DNA sequence was used in the
XX exemplification of the invention.

SQ Sequence 4470 BP; 1254 A; 991 C; 1054 G; 1171 T; 0 U; 0 Other;

Query Match 100.0%; Score 261; DB 9; Length 4470;
Best Local Similarity 100.0%; Pred. No. 1.3e-40;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGGCGCTGCTGGAGG 60
DB 366 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGGCGCTGCTGGAGG 425

QY 61 GCTGAGCCTCGCGGACGTGTTCTCTGGCGAGACGGTGCCTTTCATCAAGACCATCCGGCT 120
DB 426 GCTGAGCCTCGCGGACGTGTTCTCTGGCGAGACGGTGCCTTTCATCAAGACCATCCGGCT 485

QY 121 CQTGCGGCCAGTCTGTCCTCGGCCACCGGGAGCCGATGCCCTGAAGGGGAGCGCT 180
DB 486 CQTGCGGCCAGTCTGTCCTCGGCCACCGGGAGCCGATGCCCTGAAGGGGAGCGCT 545

QY 181 GCCCGCCGCTGCCCGGAGGAGCTGCTTCAGGCGGAGGTGGAGTACAACGGGGGCTT 240
DB 546 GCCCGCCGCTGCCCGGAGGAGCTGCTTCAGGCGGAGGTGGAGTACAACGGGGGCTT 605

QY 241 CCACCTGGCCATCGACGTGGA 261

DB 606 CCACCTGGCCATCGACGTGGA 626

RESULT 13
ADD49052
ID ADD49052 standard; DNA; 4801 BP.

XX AC ADD49052;

XX DT 15-JAN-2004 (first entry)

XX DE Human NOV6a coding sequence, SEQ ID 25.

XX KW Antidiabetic; anorectic; cardiant; hypotensive; antiarteriosclerotic;
XX KW virucide; antibacterial; fungicide; protozoacide; nootropic;
XX KW neuroprotective; antiparkinsonian; anticonvulsant; osteopathic;
XX KW antiarthritic; antiinflammatory; dermatological; antidiabetic;
XX KW antilipemic; gene therapy; NOV protein; metabolic disorder; diabetes;
XX KW obesity; viral infection; bacterial infection; fungal infection;
XX KW helminthic infection; protozoal infection; anorexia; cancer;
XX KW cardiovascular disease; hypertension; atherosclerosis;
XX KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
XX KW epilepsy; immune disorder; osteoarthritis; haematopoietic disorder;

KW inflammatory skin disorder; asthma; dyslipidemia; human; gene; ds.

XX Homo sapiens.

OS WO2003060149-A2.

XX 24-JUL-2003.

XX 06-JAN-2003; 2003WO-US000252.

XX 04-JAN-2002; 2002US-0345222P.

XX 14-JAN-2002; 2002US-0348693P.

XX 16-JAN-2002; 2002US-0349182P.

XX 17-JAN-2002; 2002US-0349733P.

XX 18-JAN-2002; 2002US-0350263P.

XX 24-JAN-2002; 2002US-0351977P.

XX 28-MAY-2002; 2002US-0383758P.

XX 05-JUN-2002; 2002US-0385969P.

XX 11-JUN-2002; 2002US-0387834P.

XX 17-JUL-2002; 2002US-0396407P.

XX 30-SEP-2002; 2002US-0415115P.

XX 03-JAN-2003; 2003US-00336603.

XX (CURA-) CURAGEN CORP.

XX Grosse WM, Alsobrook JP, Anderson DW, Burgess CE, Edinger SR;

XX Ellerman K, Furtak K, Gangolli EA, Gerlach VL, Gilbert JA;

XX Gunther E, Gorman L, Guo X, Ji W, Li L, Miller CE, Padigar M;

XX Patturajan M, Rastelli L, Macdougall JR, Mishra VS, Smithson G;

XX Spytek KA, Stone DJ, Shenoy SG, Taupier RJ, Vernet CAM, Zhong M;

XX Malyankar UM, Millet I, Kekuda R;

XX WPI; 2003-587288/55.

XX P-PSDB; ADD49053.

XX New isolated NOVX polypeptides and polynucleotides, useful for

XX preventing, diagnosing or treating NOVX-associated disorders, e.g.

XX osteoarthritis, obesity, atherosclerosis, cancer, Parkinson's disease,

XX asthma, or infections.

XX Claim 20; Page 126-127; 31lpp; English.

XX The present invention relates to novel NOV proteins and their coding

XX sequences (ADD49028-ADD49131). The proteins and coding sequences are

XX useful in the manufacture of a medicament for treating a syndrome

XX associated with a human disease, preferably a NOV-associated disorder

XX such as metabolic disorders, diabetes, obesity, infectious diseases

XX (viral, bacterial, fungal, helminthic, and protozoal), anorexia, cancer,

XX cardiovascular diseases (hypertension, atherosclerosis),

XX neurodegenerative disorders, Alzheimer's disease, Parkinson's disease,

XX epilepsy, immune disorders (osteoarthritis), hematopoietic disorders,

XX inflammatory skin disorders, asthma and various dyslipidemias. The coding

XX sequences and proteins may also be used as targets for the identification

XX of small molecules that modulate or inhibit e.g. neurogenesis, cell

XX differentiation, cell proliferation, hematopoiesis, wound healing and

XX angiogenesis, in gene therapy, in generation of antibodies that bind

XX immunospecifically to NOV substances for use in therapeutic or diagnostic

XX methods.

XX Sequence 4801 BP; 1358 A; 1037 C; 1142 G; 1264 T; 0 U; 0 Other;

XX Query Match 100.0%; Score 261; DB 9; Length 4801;

XX Best Local Similarity 100.0%; Pred. No. 1.3e-40;

XX Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1 GATCAAGTGGAGTTCGAGAGCTGCTGCAGACACGCGCGCGCTGTGGAGGG 60

XX 543 GATCAAGTGGAGTTCGAGAGCTGCTGCAGACACGCGCGCGCTGTGGAGGG 602

XX 61 GCTGAGCTGCGGACGTGTTCCTGGGCGAGACGGTCCCTTCATCAAGACCATCCGGCT 120

XX 603 GCTGAGCTGCGGACGTGTTCCTGGGCGAGACGGTCCCTTCATCAAGACCATCCGGCT 662

QY 121 CGTGGCGCCAGTCTGCTCCCTCGCCACCGGGAGCCCGATGCGCCCTGAAGGGAGCGCT 180

Db 663 CGTGGCGCCAGTCTGCTCCCTCGCCACCGGGAGCCCGATGCGCCCTGAAGGGAGCGCT 722

QY 181 GCCCGCGCTGCCCCGAGGAGCTGCGCTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 240

Db 723 GCCCGCGCTGCCCCGAGGAGCTGCGCTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 782

QY 241 CCACCTGGCCATCGAGCTGGA 261

Db 783 CCACCTGGCCATCGAGCTGGA 803

RESULT 14

ABX95681

ID ABX95681 standard; cDNA; 263 BP.

XX AC

XX ABX95681;

XX DT 30-JUN-2003 (first entry)

XX DE Rat GliTEN expressed sequence tag.

XX KW Rat; ss; EST; glioblastoma multiforme; brain cancer; GliTEN; GMB;

XX OS Rattus Norvegicus.

XX PN US2003044811-A1.

XX PD 06-MAR-2003.

XX PF 20-OCT-2001; 2001US-00051769.

XX PR 20-OCT-2000; 2000US-0242160P.

XX PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

XX PI McKinnon RD;

XX PS WPI; 2003-418995/39.

XX PT New isolated nucleic acid representing a gene product associated with

XX Glioblastoma Multiforme, designated as GliTEN, useful as a molecular

XX marker for tumor cell identification and classification.

XX Example 2; Page 4; 1lpp; English.

XX The invention relates to an isolated nucleic acid representing a gene

XX product associated with Glioblastoma Multiforme, designated as GliTEN,

XX comprising the human EST (expressed sequence tag) appearing as ABX95678,

XX or a sequence that hybridises under stringent conditions to the EST, or

XX its complement. Also included are a probe for use in identifying a

XX patient at risk for progression into the malignant phenotype, comprising

XX the nucleic acids detailed above, detecting whether a patient is at risk

XX for progression into Glioblastoma Multiforme (comprising: (a) providing a

XX sample from a patient; (b) adding the probe to the sample or performing a

XX PCR analysis using the primers appearing as ABX95682 and ABX95683; (c)

XX analysing levels of mRNA bound with the probe; and (d) treating a control

XX sample to assess the level of mRNA in a control sample, where the

XX presence of increased levels of mRNA expression in the sample in an

XX amount higher than the control sample indicates risk for progression into

XX Glioblastoma Multiforme), and a kit for use in detecting whether a

XX patient is at risk for progression into Glioblastoma Multiforme

XX (comprising the probe or primers). The probe or the kit is useful for

XX detecting whether a patient is at risk for progression into Glioblastoma

XX Multiforme (a type of brain cancer). The EST useful as molecular marker

XX for tumour cell identification and classification and for diagnosing or

XX identifying candidates at risk for progression into a malignant phenotype

XX especially in brain cancer therapy, treatment and diagnosis. The present

XX sequence is a Rat GliTEN expressed sequence tag

XX Sequence 263 BP; 41 A; 78 C; 94 G; 50 T; 0 U; 0 Other;

Query Match 79.8%; Score 208.2; DB 7; Length 263;
Best Local Similarity 87.4%; Pred. No. 1.4e-30;
Matches 228; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 1 GATCAAGGTGGAGTTCTGAGGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGAGGG 60
DB 1 GATCAAGGTGGAGTTCTGAGGAGCTGCTGCAGACCAAGACGGCGCTTTTTTTTTTTGGAGGG 60

QY 61 GCTGAGCCTGCGGGAGCTGTTCTTGGGCGAGACGGTGGCCCTTCATCAAGACCATCCGGCT 120
DB 61 GCTGAGCCTGCGGGAGCTGTTCTTGGGCGAGCTGTTCTTGGGCGACCCGCGGACCCGATGGGGACGCTCT 120

QY 121 CGTGGCGCCAGTCTGTCCTCGCCCTCGGCCACCGGGGAGCCCGATGGCCCTGAAGGGGAGCGCT 180
DB 121 GGTGGCGCCAGTCTGTCCTCGCCCTCGGCCACCGGGGAGCCCGATGGCCCTGAAGGGGAGCGCT 180

QY 181 GCCCGCCGCTGCTGCGGAGGAGCTGCTTCCGAGCGGAGGTGGAGTACAAACGGGGGCTT 240
DB 181 GCCCGCCGCTGCTGCGGAGGAGCTGCTTCCGAGCGGAGGTGGAGTACAAACGGGGGCTT 240

QY 241 CCACCTGGCCATCGACGTGGA 261
DB 241 CCACCTGGCCATCGACGTGGA 261

Search completed: July 18, 2004, 12:45:02
Job time : 438.57 secs

RESULT 15
AAD61561
ID AAD61561 standard; cDNA; 263 BP.
XX AC AAD61561;
XX DT 15-JAN-2004 (first entry)
XX DE Human GliTEN EST clone cDNA #2.
XX KW Human; glioblastoma multiforme; GBM; GliTEN; brain cancer; diagnosis;
XX KW gene therapy; expressed sequence tag; ss.
XX OS Homo sapiens.
XX FN US2003108915-A1.
XX PD 12-JUN-2003.
XX PF 20-AUG-2002; 2002US-00224624.
XX PR 20-OCT-2000; 2000US-0242160P.
XX PR 20-OCT-2001; 2001US-00051769.
XX PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
XX PI McKinnon RD;
XX DR WPI; 2003-810848/76.
XX PT Novel glioblastoma multiforme associated protein GliTEN useful for
XX treating glioblastoma multiforme and diagnosing brain cancer.
XX PS Example 2; Page 8; Opp; English.
XX CC The present invention provides novel glioblastoma multiforme (GBM)
XX CC associated protein GliTEN useful for treating glioblastoma multiforme and
XX CC diagnosing brain cancer. The invention is useful in brain cancer therapy,
XX CC treatment and diagnosis. The invention is also useful in gene therapy.
XX CC The present sequence is human GliTEN EST clone cDNA. This EST clone is
XX CC used in the exemplification of the invention
SQ Sequence 263 BP; 41 A; 78 C; 94 G; 50 T; 0 U; 0 Other;

Query Match 79.8%; Score 208.2; DB 9; Length 263;
Best Local Similarity 87.4%; Pred. No. 1.4e-30;
Matches 228; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 11:55:38 ; Search time 3237.26 Seconds
(without alignments)
2407.604 Million cell updates/sec

Title: US-10-051-769-2
Perfect score: 261
Sequence: 1 gatcaagtgaggcttcgagg.....cacctggccatcgacgtgga 261

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vit:*

21: em_gss_fun:*

22: em_gss_man:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_pbg:*

27: em_gss_vri:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	261	100.0	583	10	BF330450
2	261	100.0	3069	11	BC033019 Homo sapi
3	259	99.2	989	13	BX341552
4	227.4	87.1	458	10	BB638573

5	227.4	87.1	702	13	BU706646
6	227.4	87.1	930	13	BQ960960
7	227.4	87.1	2052	11	AK041847
8	227.2	87.0	738	29	CC567140
9	226.4	86.7	437	13	BY276266
10	203	77.8	944	13	BX345030
11	104.8	40.2	681	10	BB866050
12	102.6	39.3	730	13	BU449181
13	102.6	39.3	898	13	BU453045
14	92.6	35.5	667	13	BY734796
15	90	34.5	723	9	AW013379
16	86	33.0	618	9	AL630529
17	86	33.0	853	13	BX775059
18	80.8	31.0	464	10	BB655451
C 19	71	27.2	523	13	BU443584
C 20	62.6	24.0	1145	28	CC290954
21	52	19.9	936	14	CA475456
22	51	19.5	517	13	BX611556
23	51	19.5	682	12	BM614172
24	51	19.5	900	13	BX606255
25	50.4	19.3	558	10	BB600131
26	50.4	19.3	614	12	BM318200
C 27	50.4	19.3	803	29	CC682984
C 28	50.4	19.3	804	29	CG289528
C 29	49.4	18.9	561	28	CC162597
C 30	49.4	18.9	852	28	BZ644696
C 31	49.4	18.9	912	29	CG332228
C 32	49.2	18.9	572	9	AL876437
33	49.2	18.9	582	12	BG709715
34	49.2	18.9	598	12	BI724437
35	49.2	18.9	598	13	BU652577
36	49.2	18.9	704	14	CF865047
37	49.2	18.9	713	14	CF525078
38	49.2	18.9	1000	13	BX407619
39	49	18.8	435	9	AL588846
C 40	49	18.8	633	14	CA219654
C 41	49	18.8	639	14	CA173953
42	48.8	18.7	391	29	CG423113
43	48.8	18.7	490	12	BG320293
44	48.6	18.6	557	14	CB926331
C 45	48.6	18.6	615	14	CD453133

ALIGNMENTS

RESULT 1
BF330450/c
LOCUS BF330450 583 bp mRNA linear EST 22-NOV-2000
DEFINITION MR2-BN0364-280800-005-f07 BN0364 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF330450
VERSION BF330450.1 GI:11301198
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 583)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jorgensen,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

PubMed 10737800

CONTACT: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asmpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=WR2&t2=WR2-BN0364-280800-005-f07&t3=2000-08-28&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 10
High quality sequence stop: 581.

FEATURES

source
1..583
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="BN0364"
/note="Torgan: breast normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 100.0%; Score 261; DB 10; Length 583;
Best Local Similarity 100.0%; Pred. No. 1.9e-32;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCCGGCGCCCTGCTGGAGGG 60
DB 482 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCCGGCGCCCTGCTGGAGGG 423
QY 61 GCTGAGCCTGGCGGACGCTGTTCTCTGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
DB 422 GCTGAGCCTGGCGGACGCTGTTCTCTGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 363
QY 121 CGTGGGCGAGTGTGTCCTCGCCGACACGGGGAGCCCGATGCGCCCTGAAGGGAGGCGCT 180
DB 362 CGTGGGCGAGTGTGTCCTCGCCGACACGGGGAGCCCGATGCGCCCTGAAGGGAGGCGCT 303
QY 181 GCCCGCGCCCTGCCCGGAGGAGTGGCTTCGAGCGGAGGTGGAGTACACGGGGGCTT 240
DB 302 GCCCGCGCCCTGCCCGGAGGAGTGGCTTCGAGCGGAGGTGGAGTACACGGGGGCTT 243
QY 241 CCACCTGGCCATCGACGTGA 261
DB 242 CCACCTGGCCATCGACGTGA 222

RESULT 2

BC033019
LOCUS BC033019 3069 bp mRNA linear HTC 19-NOV-2003
DEFINITION Homo sapiens hypothetical protein LOC118987, mRNA (cDNA clone IMAGE:4821752), containing frame-shift errors.

ACCESSION

BC033019

VERSION

BC033019.1 GI:21542573

KEYWORDS

HTC.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3069)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J.M., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,

Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, D.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzyzanski, M.I., Skalska, U., Smalish, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257

12477932

2 (bases 1 to 3069)

Strausberg, R.

Direct Submission

JOURNAL

Submitted (21-JUN-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: http://mgc.nci.nih.gov

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

cDNA library Preparation: Michael J. Brownstein (NHGRI) & Shiraki

Toshiyuki and Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Institute for Systems Biology

http://www.systemsbio.org

contact: amadansystemsbiology.org

Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha

Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Series: IPAK Plate: 33 Row: b Column: 4

This clone was selected for full length sequencing because it

passed the following selection criteria: Genomescan gene prediction

This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers
source
1..3069
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4821752"
/tissue_type="Testis"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"

ORIGIN

Query Match 100.0%; Score 261; DB 11; Length 3069;
Best Local Similarity 100.0%; Pred. No. 2.6e-32;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCCGGCGCCCTGCTGGAGGG 60
DB 610 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCCGGCGCCCTGCTGGAGGG 669
QY 61 GCTGAGCCTGGCGGACGCTGTTCTCTGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
DB 670 GCTGAGCCTGGCGGACGCTGTTCTCTGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 729
QY 121 CGTGGGCGAGTGTGTCCTCGCCGACACGGGGAGCCCGATGCGCCCTGAAGGGAGGCGCT 180
DB 730 CGTGGGCGAGTGTGTCCTCGCCGACACGGGGAGCCCGATGCGCCCTGAAGGGAGGCGCT 789
QY 181 GCCCGCGCCCTGCCCGGAGGAGTGGCTTCGAGCGGAGGTGGAGTACACGGGGGCTT 240
DB 790 GCCCGCGCCCTGCCCGGAGGAGTGGCTTCGAGCGGAGGTGGAGTACACGGGGGCTT 849

```

QY      241 CCACCTGGCCATCGAGTGA 261
      |||
Db      850 CCACCTGGCCATCGAGTGA 870
      |||

RESULT 3
LOCUS   BX341552
DEFINITION BX341552 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
          cDNA clone CS0DK002YA12 5-PRIME, mRNA sequence.
ACCESSION BX341552
VERSION   BX341552.1 GI:30339998
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Li W.B., Gruber C., Jessee J. and Polayes D.
TITLE    Full-length cDNA libraries and normalization
JOURNAL  Unpublished (2001)
COMMENT  Contact: Genoscope
          Genoscope - Centre National de Sequencage
          BP 191 91006 EVRY cedex - France
          Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
          Library was constructed by Life Technologies, a division of
          Invitrogen. This sequence belongs to sequence cluster 5483.r For
          more information about this cluster, see
          http://www.genoscope.cns.fr/
          cgi-bin/cluster.cgi?seq=CS0DK002BA06QP1&cluster=5483.r. Contact :
          Feng Liang Email : fliang@lifetech.com URL :
          http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
          Faraday Avenue Genoscope sequence ID : CS0DK002BA06QP1.
          Location/Qualifiers
            1..989
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="CS0DK002YA12"
              /cell_type="HELA CELLS COT 25-NORMALIZED"
              /cell_line="HELA"
              /clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
              /note="1st strand cDNA was primed with a NotI-oligo(dT)
              primer. Five prime end enriched, double-strand cDNA was
              digested with Not I and cloned into the Not I and EcoR V
              sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match          99.2%; Score 259; DB 13; Length 989;
Best Local Similarity 99.2%; Pred. No. 4.3e-32;
Matches 259; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 GATCAAGTGGAGTTCGAGAGCTGCTGCAGACCAAGACGGCGCGCGCTGCTGGAGGG 60
      |||
Db      633 GATCAAGTGGAGTTCGAGAGCTGCTGCAGACCAAGACGGCGCGCGCTGCTGGAGGG 692
      |||

QY      61 GCTGAGCCTGCGGAGCGTGTTCCTGGGCGAGACGCTGCCCTTCATCAAGACCATCCGGCT 120
      |||
Db      693 GCTGAGCCTGCGGAGCGTGTTCCTGGGCGAGACGCTGCCCTTCATCAAGACCATCCGGCT 752
      |||

QY      121 CFTGCGGCCAGTGTGCTCCCTCGCCACCGGGAGCCGATGCCCTGAAGGGGAGCGCT 180
      |||
Db      753 CFTGCGGCCAGTGTGCTCCCTCGCCACCGGGAGCCGATGCCCTGAAGGGGAGCGCT 812
      |||

QY      181 GCCCGCGCGCTGCCCGGAGGAGTGGCTTCGAGGCGGAGGTGAGTACAAACGGGGGCGCTT 240
      |||
Db      813 GCCCGCGCGCTGCCCGGAGGAGTGGCTTCGAGGCGGAGGTGAGTACAAACGGGGGCGCTT 872
      |||

QY      241 CCACCTGGCCATCGAGTGA 261
      |||
Db      873 CCACCTGGCCATCGAGTGA 893
      |||

```

```

RESULT 4
BB638573
LOCUS   BB638573
DEFINITION BB638573 RIKEN full-length enriched, 3 days neonate thymus Mus
          musculus cDNA clone A630041P07 5', mRNA sequence.
ACCESSION BB638573.1 GI:15401196
VERSION   BB638573
KEYWORDS EST.
SOURCE   Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 458)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hangaki,T.,
          Hara,A., Hiranoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
          Komno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
          Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
          Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
          Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
          Takeda,Y., Tanaka,T., Toyota,T., Muramatsu,M. and Hayashizaki,Y.
          RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
          Unpublished (2001)
          Contact: Yoshihide Hayashizaki
          Laboratory for Genome Exploration Research Group, RIKEN Genomic
          Sciences Center (GSC), Yokohama Institute
          The Institute of Physical and Chemical Research (RIKEN)
          1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
          Tel: 81-45-503-9222
          Fax: 81-45-503-9216
          Email: genome-res@gsc.riken.go.jp,
          URL:http://genome.gsc.riken.go.jp/
          Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
          Itoh,M., Komno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
          Normalization and subtraction of cap-trapper-selected cDNAs to
          prepare full-length cDNA libraries for rapid discovery of new
          genes. Genome Res. 10 (10), 1617-1630 (2000)
          wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
          Watanuki,M., Itoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
          Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
          and Hayashizaki,Y.
          RIKEN integrated sequence analysis (RISA) system--384-format
          sequencing pipeline with 384 multicapillary sequencer. Genome Res.
          10 (11), 1757-1771 (2000)
          Komno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
          Sugahara,Y. and Hayashizaki,Y.
          Computer-based methods for the mouse full-length cDNA
          encyclopedia: real-time sequence clustering for construction of a
          nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
          Yananaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A.,
          Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K.,
          Arakawa,T., Ishii,Y. and Hayashizaki,Y.
          Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.
          Func. Genomics 2 pre, L72-L86 (2001)
          Please visit our web site (http://genome.gsc.riken.go.jp) for
          further details.
          e mouse tissues.
            Location/Qualifiers
              1..458
                /organism="Mus musculus"
                /mol_type="mRNA"
                /db_xref="taxon:10090"
                /clone="A630041P07"
                /tissue_type="thymus"
                /dev_stage="3 days neonate"
                /lab_host="DH10B"
                /clone_lib="RIKEN full-length enriched, 3 days neonate
                thymus"
                /note="Site_1: Sall; Site_2: BamHI; cDNA library was
                prepared and sequenced in Mouse Genome Encyclopedia
                Project of Genome Exploration Research Group in Riken
                Genomic Sciences Center and Genome Science Laboratory in
                RIKEN Division of Experimental Animal Research in Riken
                contributed to prepare mouse tissues. 1st strand cDNA was

```

```

FEATURES
source

```

primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGATCTCGAGTTAATTAATATCCCTCCCTCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."

ORIGIN

Query Match 87.1%; Score 227.4; DB 10; Length 458;
Best Local Similarity 92.0%; Pred. No. 4.3e-27;
Matches 240; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 1 GATCAAGGTGGAGTTTCGAGGAGCTCTGCAGACCAAGACGGCGCGGCGCTTGCCTGGAGGG 60
DB 6 GATCAAGGTGGAGTTTCGAGGAGCTCTGCAGACCAAGACGGCGCGGCGCTTGCCTGGAGGG 65
QY 61 GCTGAGCTGGGGGACGTTCTCTGGGCGAGCGGTGCTTCATCAAGACCATCCGGCT 120
DB 66 GCTGAGTCTGGCGGACGTTCTCTGGGCGACACGGTGCCTTCATCAAGACCATCCGGCT 125
QY 121 CGTGGCGGACGTTGCTCGCCCTCGGACCGGGGACCGGATGGCCCTGAGGGGAGCGCT 180
DB 126 GGTGGGCGGCGGTTGGTGGCTCTGGGACCGGGGAGCGGACCGCGGAGCGCGCT 185
QY 181 GCCCGCCGCTGCCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT 240
DB 186 GCCCGCCACCTGCCCGAGGAGCTGGCCCTTGAACGGGAGGTGGAGTACAAACGGGGGCTT 245
QY 241 CCACCTGGCCATCGACGTGGA 261
DB 246 CCACCTGGCCATCGACGTGGA 266

RESULT 5

BU706646
LOCUS
DEFINITION
UI-M-FOO-cad-p-11-0-UI_r1 NIH BMAP_F00 Mus musculus cDNA clone
IMAGE:6409978 5', mRNA sequence.
ACCESSION
BU706646
VERSION
BU706646.1 GI:23637322
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution by: Dr. M. Bento Soares, University of Iowa
Clone Distribution by: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: pYX-5.

FEATURES

Location/Qualifiers
1..702
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6409978"

/tissue type="whole brain"
/dev_stage="embryo 12.5dpc"
/lab_host="DHI0B (T1 phage resistant)"
/clone_lib="NIH BMAP_F00"

/note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is TGAGAGAGCC. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."

ORIGIN

Query Match 87.1%; Score 227.4; DB 13; Length 702;
Best Local Similarity 92.0%; Pred. No. 4.7e-27;
Matches 240; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 1 GATCAAGGTGGAGTTTCGAGGAGCTCTGCAGACCAAGACGGCGCGGCGCTTGCCTGGAGGG 60
DB 310 GATCAAGGTGGAGTTTCGAGGAGCTCTGCAGACCAAGACGGCGCGGCGCTTGCCTGGAGGG 369
QY 61 GCTGAGCTGGGGGACGTTCTCTGGGCGAGCGGTGCTTCATCAAGACCATCCGGCT 120
DB 370 GCTGAGTCTGGCGGACGTTCTCTGGGCGACACGGTGCCTTCATCAAGACCATCCGGCT 429
QY 121 CGTGGCGGACGTTGCTCGCCCTCGGACCGGGGACCGGATGGCCCTGAGGGGAGCGCT 180
DB 430 GGTGGGCGGCGGTTGGTGGCTCTGGGACCGGGGAGCGGACCGCGGAGCGCGCT 489
QY 181 GCCCGCCGCTGCCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT 240
DB 490 GCCCGCCACCTGCCCGAGGAGCTGGCCCTTGAACGGGAGGTGGAGTACAAACGGGGGCTT 549
QY 241 CCACCTGGCCATCGACGTGGA 261
DB 550 CCACCTGGCCATCGACGTGGA 570

RESULT 6

BU960960
LOCUS
DEFINITION
AGENCOURT 8955743 NCI CGAP_Mam2 Mus musculus cDNA clone
IMAGE:6439792 5', mRNA sequence.
ACCESSION
BU960960
VERSION
BU960960.1 GI:22376438
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution by: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAMI3960 row: f column: 17
High quality sequence stop: 603.

FEATURES		Location/Qualifiers	
source		1..930	
		/organism="Mus musculus"	
		/mol_type="mRNA"	
		/strain="FVB/N-3"	
		/db_xref="taxon:10090"	
		/clone="IMAGE:6439792"	
		/tissue_type="tumor, biopsy sample"	
		/dev_stage="5 months"	
		/lab_host="DH10B"	
		/clone_lib="NCI CGAP Mam2"	
		/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"	
ORIGIN			
Query Match		87.1%; Score 227.4; DB 13; Length 930;	
Best Local Similarity		92.0%; Pred. No. 4.9e-27;	
Matches 240; Conservative		0; Mismatches 21; Indels 0; Gaps 0;	
QY	1	GATCAAGTGGAGTTCGAGGAGTGTGCGAGACCAAGACGGCGCGCGCTGCTGGAGGG	60
Db	335	GATCAAGTGGAGTTCGAGGAGTGTGCGAGACCAAGACGGCGCGCGCTGCTGGAGGG	394
QY	61	GCTGAGCCTCGGGAGCGTGTCTCTGGCGGAGACGGTTCCTTCATCAAGACCATCCGGCT	120
Db	395	GCTGAGTCTCGCGAGCGTGTCTCTGGCGGAGACGGTTCCTTCATCAAGACCATCCGGCT	454
QY	121	CTGCGGCGCAGTGTGCTCCCTCGCCACCGGGAGCCCGATGGCCCTGAAGGGAGGCGCT	180
Db	455	GGTGCGGCCCGTGTGCTCGGCGCAACCGCGAGCCGACCCCGACGGGAGCGCGCT	514
QY	181	GCCCGCGCTGCGCCGAGGAGTGTGCGAGCGGAGTGGAGTACAAACGGGGCGCTT	240
Db	515	GCCCGCCACCTGCGCGGAGGAGTGTGCGAGCGGAGTGGAGTACAAACGGGGCGTTT	574
QY	241	CCACTTGGCCATCGAGCTGGA	261
Db	575	CCACTTGGCCATCGAGCTGGA	595
RESULT 7			
AK041847			
LOCUS			
DEFINITION			
Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched library, clone:A630041P07 product:hypothetical PDZ domain (also known as DHR or GLGF) containing protein, full insert sequence.			
AK041847			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Mus musculus			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
1			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			

```
Matches 240; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 GATCAAGTGGAGTTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGGCGCTCTGAGAGG 60
DB |||||
10 GATCAAGTGGAGTTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGGCGCTCTGAGAGG 69
DB |||||

QY 61 GCTGAGCCTGCGGAGCGTCTCTCGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
DB |||||
70 GCTGAGTCTGCGGAGCGTCTCTCGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 129
DB |||||

QY 121 COTCGGCGAGTCGTGCTCCCTCGGCCACCGGGAGCCCGATGCCCTCGAAGGGGAGCGCT 180
DB |||||
130 GGTGCGGCGCGTGTGGCTCGGCGACCGGCGAGCCCGACGACCGGCGACGCGCT 189
DB |||||

QY 181 GCCCGCGCTGCCCGGAGGAGCTGGCTTCAGGCGGAGTGGAGTACAAAGGGGGCTT 240
DB |||||
190 GCCCGCGGAGCTGCCCGGAGGAGCTGGCTTTGAAGCGGAGTGGAGTACAAAGCGGGTTT 249
DB |||||

QY 241 CCACCTGGCCATCGACGTGGA 261
DB |||||
250 CCACCTGGCCATCGACGTGGA 270
DB |||||

CC567140 738 bp DNA linear GSS 18-JUN-2003
CH240_441017.TARBAC13P2 CHORI-240 Bos taurus genomic clone
CH240_441017, genomic survey sequence.
CC567140
GSS.
CC567140.1 GI:31900496
Bos taurus (cow)
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 738)
Holt,R., Cloutier,A., Lee,D., Girn,N., Olson,T., Mayo,M.,
Tsai,M., Cloutier,A., Kirkpatrick,R., Liu,J., Guin,R., Chan,A., Chiu,R.,
Butterfield,Y., Mathewson,C., Wye,N., Masson,A., Brown-John,M., Jones,S.,
Schein,J., Marra,M., de Jong,P., Keeler,J.W. and Kappes,S.M.
Bovine BAC End Sequences from Library CHORI-240, PLATES 399 to 478
Unpublished (2003)
Other GSSs: CH240_441017.T7
Contact: Rob Holt
Sequencing
The British Columbia Cancer Agency Genome Science Centre
600 W. 10th Ave, Vancouver, British Columbia, Canada V5Z 4E6
Tel: 604-877-6085
Fax: 604-877-6276
Email: rholt@cgsc.ca
Clones are derived from the bovine BAC library CHORI-240
(http://www.chori.org/bacpac/bovine240.htm). For BAC library
availability, please contact Pieter de Jong (pjejong@mail.choi.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/ordering/information.htm). This work
was undertaken as part of the International Bovine BAC Mapping
Consortium (IBMC) by CSIRO Livestock Industries, Australia and the
British Columbia Genome Sciences Centre, Canada.
Plate: 441 row: 0 column: 17
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..738
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="breed: Hereford"
/db_xref="taxon:9913"
/clone="CH240_441017"
/sex="Male"
/cell_type="Blood"
/clone_lib="CHORI-240"

FEATURES
source
1..738
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="breed: Hereford"
/db_xref="taxon:9913"
/clone="CH240_441017"
/sex="Male"
/cell_type="Blood"
/clone_lib="CHORI-240"

/note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC
library (Male) produced by Pieter de Jong"
```

ORIGIN

```
Query Match 87.0%; Score 227.2; DB 29; Length 738;
Best Local Similarity 93.9%; Pred. No. 5.1e-27;
Matches 248; Conservative 0; Mismatches 13; Indels 3; Gaps 1;

QY 1 GATCAAGTGGAGTTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGGCGCTCTGAGAGG 60
DB |||||
343 GATCAAGTGGAGTTTCGAGGAGCTGCTGCAGACCAAGACGGCTGGCGGCTCTGAGAGG 402
DB |||||

QY 61 GCTGAGCCTGCGGAGCGTCTCTCGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
DB |||||
403 GCTGAGCCTGCGGAGCGTCTCTCGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 462
DB |||||

QY 121 COTCGGCGAGTCGTGCTCCCTCGGCCACCGGGAGCCCGATGCCCTCGAAGGGGAGCGCT 180
DB |||||
463 COTCGGCGAGTCGTGCTCCCTCGGCCACCGGGAGCCCGATGCCCTCGAAGGGGAGCGCT 522
DB |||||

QY 181 ---GCCCGCGCTGCCCGGAGGAGCTGGCTTCGAGGCGGAGTGGAGTACAAAGGGGG 237
DB |||||
523 GCCCGCGCTGCCCGGAGGAGCTGGCTTCGAGGCGGAGTGGAGTACAAAGGGGG 582
DB |||||

QY 238 CTTCCACCTGGCCATCGACGTGGA 261
DB |||||
583 CTTCCACCTGGCCATCGACGTGGA 606
DB |||||
```

RESULT 9

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..738

/organism="Bos taurus"

/mol_type="genomic DNA"

/strain="breed: Hereford"

/db_xref="taxon:9913"

/clone="CH240_441017"

/sex="Male"

/cell_type="Blood"

/clone_lib="CHORI-240"

/analysis="functional annotation"

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

EST 11-DEC-2002

linear

musculus

Mus musculus

Mus musculus

Mus musculus

Mus musculus

Mus musculus

Mus musculus

Mus musculus

Mus musculus

Mus musculus

Mus musculus

Mus musculus

Mus musculus

Mus musculus

MEDLINE
PUBMED
COMMENT

22354683
12466851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and
Hayashizaki, Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
non-redundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Michela Paggiolini and Takao K. Hensch (
Laboratory for Neuronal Circuit Development Brain Science Institute
RIKEN 2-1 Hirosewa, Wako-shi, Saitama 351-0198 Japan) whose
assistance we gratefully acknowledge. Please visit our web site
(http://genome.gsc.riken.go.jp) for further details.
Location/Qualifiers
1. .437
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K430318E15"
/tissue type="visual cortex"
/clone_lib="RIKEN full-length enriched, visual cortex"

ORIGIN

Query Match 86.7%; Score 226.4; DB 13; Length 437;
Best Local Similarity 91.6%; Pred. No. 6.2e-27;
Matches 239; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 1 GATCAAGTGGAGTTCGAGGAGTCTGCGAGACCAAGACGCGCGCGCTGCTGGAGGG 60
|||||
DB 81 GATCAAGTGGAGTTCGAGGAGTCTGCGAGACCAAGACGCGCGCGCTGCTGGAGGG 140
|||||
QY 61 GCTGAGCTCGGACGCTGTTCTTGGGCGAGACGCGTCCCTTCATCAAGACCATCCGCGCT 120
|||||
DB 141 GCTGAGTCTGGCGACGCTGTTCTTGTGNGGACACGCTGCCCTTCATCAAGACCATCCGCGCT 200
|||||
QY 121 CGTGGCGCCAGTCTGCTCGGCCACCGGGAGCCGATGCGCCTGAAGGGGAGGCGCT 180
|||||
DB 201 GGTGGGCGCCGCTGGTGGCTCGGGCACCGGCGAGCCCGACGACCCCGGCGGACGCGCT 260
|||||
QY 181 GCCCGCGCCTGCCCGAGGAGTGGCTTCGAGGCGGAGTGAGTACACGCGGCGCTT 240
|||||
DB 261 GCCCGCACCTGCCCGAGGAGTGGCTTCGAGGCGGAGTGAGTACACGCGGCGCTT 320
|||||
QY 241 CCACCTGGCCATCGACGTGGA 261
|||||
DB 321 CCACCTGGCCATCGACGTGGA 341
|||||

RESULT 10
BX345030
LOCUS
DEFINITION
BX345030 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
cDNA clone CS0DK002YA12 5-PRIME, mRNA sequence.
ACCESSION
BX345030
VERSION
BX345030.1
KEYWORDS
GT:30313335
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
1 (bases 1 to 944)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
JOURNAL
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5483.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS2BAX16ZB10_AX272D11_1&cluster=5483.r.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS2BAX16ZB10_AX272D11_1.
Location/Qualifiers
1. .944
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK002YA12"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (GT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES
source

Query Match 77.8%; Score 203; DB 13; Length 944;
Best Local Similarity 95.8%; Pred. No. 4e-23;
Matches 250; Conservative 0; Mismatches 7; Indels 4; Gaps 4;

QY 1 GATCAAGTGGAGTTCGAGGAGTCTGCGAGACCAAGACGCGCGCGCTGCTGGAGGG 60
|||||
DB 641 GATCAAGTGGAGTTCGAGGAGTCTGCGAGACCAAGACGCGCGCGCTGCTGGAGGG 700
|||||
QY 61 GCTGAGCTCGGCGACGCTGTTCTTGGGCGAGACGCTGCCCTTCATCAAGACCATCCGCGCT 120
|||||
DB 701 GCTGAGCTCGGCGACGCTGTTCTTGGGCGAGACGCTGCCCTTCATCAAGACCATCCGCGCT 760
|||||
QY 121 CGTGGCGCCAGTCTGCTCGGCCACCGGGAGCCGATGCGCCTGAAGGGGAGGCGCT 180
|||||
DB 761 CGTGGCGCCAGTCTGCTCGGCCACCGGGAGCCGATGG-CCTGAGGGGAGGCGCT 819
|||||
QY 181 GCCCGCGCCTGCCCGAGGAGTGGCTTCGAGGCGGAGTGAGTACACGCGGCGCTT 240
|||||
DB 820 GCCCGCGCCTGCCCGAGGAGTGGCTTCGAGGCGGAGTGAGTACACGCGGCGCTT 876
|||||
QY 241 CCACCTGGCCATCGACGTGGA 261
|||||
DB 877 CCACCTGGCCATCGACGTGGA 897
|||||

ORIGIN

Query Match 77.8%; Score 203; DB 13; Length 944;
Best Local Similarity 95.8%; Pred. No. 4e-23;
Matches 250; Conservative 0; Mismatches 7; Indels 4; Gaps 4;

QY 1 GATCAAGTGGAGTTCGAGGAGTCTGCGAGACCAAGACGCGCGCGCTGCTGGAGGG 60
|||||
DB 641 GATCAAGTGGAGTTCGAGGAGTCTGCGAGACCAAGACGCGCGCGCTGCTGGAGGG 700
|||||
QY 61 GCTGAGCTCGGCGACGCTGTTCTTGGGCGAGACGCTGCCCTTCATCAAGACCATCCGCGCT 120
|||||
DB 701 GCTGAGCTCGGCGACGCTGTTCTTGGGCGAGACGCTGCCCTTCATCAAGACCATCCGCGCT 760
|||||
QY 121 CGTGGCGCCAGTCTGCTCGGCCACCGGGAGCCGATGCGCCTGAAGGGGAGGCGCT 180
|||||
DB 761 CGTGGCGCCAGTCTGCTCGGCCACCGGGAGCCGATGG-CCTGAGGGGAGGCGCT 819
|||||
QY 181 GCCCGCGCCTGCCCGAGGAGTGGCTTCGAGGCGGAGTGAGTACACGCGGCGCTT 240
|||||
DB 820 GCCCGCGCCTGCCCGAGGAGTGGCTTCGAGGCGGAGTGAGTACACGCGGCGCTT 876
|||||
QY 241 CCACCTGGCCATCGACGTGGA 261
|||||
DB 877 CCACCTGGCCATCGACGTGGA 897
|||||

RESULT 11
BB866050
LOCUS
DEFINITION
BB866050 RIKEN full-length enriched, CRL-1751 WEHI 164 cDNA Mus


```

QY 181 GCCCGCGCCTGCCCCAGGAGCTGGCGCTTCGAGCGGAGGTGAGTACAAACGGGGGCTT 240
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 315 --AGGAGGGTGGCCCGAGGAGCTGGCGCTTCGAGGTGAGTACAAACGGCGGCTT 372
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 241 CCACCTGGCCATCGACCTGGA 261
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 373 CCACCTGGCCATCGACCGAGA 393
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
BU453045
LOCUS 898 bp mRNA linear EST 29-NOV-2002
DEFINITION 603219082F1 CSEQRN14 Gallus gallus cDNA clone CHEST210C22 5', mRNA
sequence.
ACCESSION BU453045
VERSION BU453045.1 GI:25942356
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 898)
AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES
source
1. 898
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Layer"
/db_xref="taxon:9031"
/clones="CHEST210C22"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQRN14"
/note="Organ: ovary; Vector: pBluescript II KS(+); Site_1:
EcORI; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunted, ligated to NotI adapters, digested with
EcORI, size-selected, and cloned into the NotI and EcORI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."

ORIGIN
Query Match 39.3%; Score 102.6; DB 13; Length 898;
Best Local Similarity 68.2%; Pred. No. 4.9e-07;
Matches 178; Conservative 0; Mismatches 44; Indels 39; Gaps 1;

QY 1 GATCAAGTGGAGTTCGAGAGCTGCTGCAGACCAAGACGCCCGGCGCTTGTGGAGGG 60
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 172 GATCAAGTGGAGTTCGAGAGCTGCTGCAGACCAAAATGACGGGAAGGTGTTGGAGGG 231
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 GCTGAGCCTCGGGACGTGTTCTCTGGGCGAGACGCTGCCCTTCATCAAGACCATCCGGCT 120
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db 232 GCTGAGGCTCCGGACGTGTATCTGGGCAATGTCTCCCGTCTTCAAGCGCTCCGGCT 291
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 121 CGTGGCGCCAGTGTGTCCTCGGCCACCGGGGAGCCCGATGGCCCTGAAAGGGAGGCGCT 180
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 292 CATCCGGCCGTGTGTGCAGTG----- 314
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 181 GCCCGCGCCTGCCCCAGGAGCTGGCGCTTCGAGCGGAGGTGAGTACAAACGGGGGCTT 240
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 315 --AGGAGGGTGGCCCGAGGAGCTGGCGCTTCGAGGTGAGTACAAACGGCGGCTT 372
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 241 CCACCTGGCCATCGACCTGGA 261
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 373 CCACCTGGCCATCGACCGAGA 393
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
BU734796
LOCUS 667 bp mRNA linear EST 17-DEC-2002
DEFINITION BU734796 RIKEN full-length enriched, mammary gland RCB-0526
Yyg-MC(A) cDNA Mus musculus cDNA clone G830026K01 5', mRNA
sequence.
ACCESSION BU734796
VERSION BU734796.1 GI:27147923
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 667)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaoka,I.,
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusci,V.,
Ciothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
Kawaji,H., Kawasawa,Y., Kedziarski,R.M., King,B.L., Konagaya,A.,
Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
Numata,K., Okido,T., Favan,W.J., Pertea,G., Pesole,G.,
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
Shingawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
Rogers,J., Birney,E. and Hayashizaki,Y.
ANALYSIS Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
NATURE 420, 563-573 (2002)
JOURNAL 22354683
MEDLINE 1246851
PUBMED
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp/
URL:http://genome.gsc.riken.go.jp/
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P.,
Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F.,

```

Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Nunazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers

```
1..667
/organism="Mus musculus"
/mol_type="mrna"
/db_xref="taxon:10090"
/clone="G830026K01"
/tissue_type="mammary gland"
/cell_line="RCB-0526 JY9-MC(A)"
/clone_lib="RIKEN full-length enriched, mammary gland
RCB-0526 JY9-MC(A) cDNA"
```

ORIGIN

```
Query Match      35.5%; Score 92.6; DB 13; Length 667;
Best Local Similarity 95.5%; Pred. No. 1.8e-05;
Matches 106; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTCTGAGGG 60
   |||||
Db 556 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTCTGAGGG 615
   |||||

QY 61 GCTGAGCTCGCGGACGTCTCTCGGAGAGCGGTGCCCTTCATCAAGAC 111
   |||||
Db 616 GCTGAGCTCGCGGACGTCTCTCGGAGAGCGGTGCCCTTCATCAAGAC 665
   |||||
```

RESULT 15

AW013379

LOCUS

AW013379 723 bp mRNA linear EST 10-SEP-1999
sp042ks Winter flounder spleen Pseudopleuronectes americanus cDNA
clone sp042ks 5' similar to C53B4.4 [Caenorhabditis elegans], mRNA
sequence.

ACCESSION

AW013379

VERSION

AW013379.1 GI:5862157

KEYWORDS

EST.

SOURCE

Pseudopleuronectes americanus (winter flounder)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;

Pleuronectoidei; Pleuronectidae; Pseudopleuronectes.

1 (bases 1 to 723)

Douglas, S.E., Gallant, J.W., Bullerwell, C.E., Wolff, C.,

Munholland, J. and Reith, M.E.

Winter flounder expressed sequence tags: Establishment of an EST

database and identification of novel fish genes

Marine Biotechnology (1999) In press

Contact: Reith M

Marine Biology

NRC Institute for Marine Biosciences

1411 Oxford St., Halifax, Nova Scotia, B3H 3Z1, Canada
Tel: (902) 426-8276
Fax: (902) 426-9413
Email: michael.reith@nrc.ca
Seq primer: M13 Forward.

FEATURES

Source

```
1..723
/organism="Pseudopleuronectes americanus"
/mol_type="mrna"
/db_xref="taxon:8265"
/clone="sp042ks"
/sex="female"
/dev stage="adult"
/clone_lib="Winter flounder spleen"
/note="Organ: spleen"
```

ORIGIN

```
Query Match      34.5%; Score 90; DB 9; Length 723;
Best Local Similarity 69.1%; Pred. No. 4.9e-05;
Matches 123; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTCTGAGGG 60
   |||||
Db 482 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTCTGAGGG 541
   |||||

QY 61 GCTGAGCTCGCGGAGGTCTCTCGGAGAGCGGTGCCCTTCATCAAGACCATCCGGCT 120
   |||||
Db 542 GCTGAGCTCGAGAGACATCTCCCTGGGCAATTCTCTGCCCGCTTTCAGAGCGCCACCT 601
   |||||

QY 121 CGTGGCGGCAGTCTGTCGCCCTCGGCCACCGGGGAGCGGATGCCCTGAAGGGGAGGCG 178
   |||||
Db 602 CATGAAGCGCGTGCACCTGAACGAAGACGGCATGCCGAGAACTCAACTTCGAGGTG 659
   |||||
```

Search completed: July 18, 2004, 14:36:32
Job time : 3242.26 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 14:38:35 ; Search time 265 Seconds
(without alignments)
4804.984 Million cell updates/sec

Title: US-10-051-769-2

Perfect score: 261

Sequence: 1 gatcaagtgagtgagg.....cacctggccatcgacgtgga 261

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 3190992 seqs, 2439331697 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6381984

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	261	100.0	261	US-10-051-769-2	Sequence 2, Appli
2	261	100.0	261	US-10-224-624-2	Sequence 2, Appli
3	261	100.0	873	US-10-051-769-3	Sequence 3, Appli
4	261	100.0	873	US-10-224-624-3	Sequence 3, Appli
5	261	100.0	960	US-10-051-769-1	Sequence 1, Appli
6	261	100.0	960	US-10-224-624-1	Sequence 1, Appli
7	261	100.0	2639	US-10-104-047-684	Sequence 684, App
8	261	100.0	3465	US-10-224-624-9	Sequence 9, Appli
9	261	100.0	3649	US-10-112-944-63	Sequence 63, Appli
10	261	100.0	3832	US-10-224-624-7	Sequence 7, Appli
11	261	100.0	4470	US-10-276-774-950	Sequence 950, App
12	261	100.0	4801	US-10-336-603A-25	Sequence 25, Appl
13	47	18.0	3847	US-10-250-823-11	Sequence 11, Appl
14	43	16.5	263	US-10-051-769-4	Sequence 4, Appli

15	43	16.5	263	15	US-10-224-624-4	Sequence 4, Appli
16	22	8.4	22	15	US-10-051-769-5	Sequence 5, Appli
17	22	8.4	22	15	US-10-224-624-5	Sequence 5, Appli
18	21	8.0	1425	8	US-08-900-220C-6	Sequence 6, Appli
19	21	8.0	1425	8	US-08-954-771-6	Sequence 6, Appli
20	21	8.0	1425	8	US-08-462-386D-6	Sequence 6, Appli
21	21	8.0	1425	9	US-09-021-660A-32	Sequence 32, Appli
22	21	8.0	1425	9	US-09-151-999-6	Sequence 6, Appli
23	21	8.0	1425	10	US-09-883-848A-6	Sequence 6, Appli
24	21	8.0	1425	10	US-09-187-387-6	Sequence 6, Appli
25	21	8.0	1425	10	US-09-827-110-6	Sequence 6, Appli
26	21	8.0	1425	10	US-09-845-025C-6	Sequence 6, Appli
27	21	8.0	1425	10	US-09-451-939-6	Sequence 6, Appli
28	21	8.0	1425	10	US-09-238-243-6	Sequence 6, Appli
29	21	8.0	1425	10	US-09-736-476-6	Sequence 6, Appli
30	21	8.0	1425	13	US-10-244-095A-6	Sequence 6, Appli
31	21	8.0	1425	16	US-10-294-036-6	Sequence 6, Appli
32	21	8.0	1425	17	US-10-652-298A-6	Sequence 6, Appli
33	21	8.0	1575	12	US-10-456-954-1	Sequence 1, Appli
34	21	8.0	1575	17	US-10-414-267-1	Sequence 1, Appli
35	21	8.0	1576	10	US-09-733-634-13	Sequence 13, Appli
36	21	8.0	1576	15	US-10-001-844-3	Sequence 3, Appli
37	21	8.0	2274	15	US-10-228-931-3	Sequence 3, Appli
38	19	7.3	348	15	US-10-241-009-44	Sequence 44, Appli
39	19	7.3	348	15	US-10-241-009-45	Sequence 45, Appli
40	19	7.3	348	15	US-10-190-434B-44	Sequence 44, Appli
41	19	7.3	348	15	US-10-190-434B-45	Sequence 45, Appli
42	19	7.3	348	15	US-10-190-305A-57	Sequence 57, Appli
43	19	7.3	348	15	US-10-190-305A-58	Sequence 58, Appli
44	19	7.3	550	17	US-10-437-963-25242	Sequence 25242, A
45	19	7.3	1281	15	US-10-241-009-51	Sequence 51, Appli

ALIGNMENTS

RESULT 1

US-10-051-769-2
; Sequence 2, Application US/10051769
; Publication No. US2003004481A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randy D.
; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION
; FILE REFERENCE: 268/260 (RMW-00-37)
; CURRENT APPLICATION NUMBER: US/10/051,769
; CURRENT FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,160
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 261
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-051-769-2

Query Match 100.0%; Score 261; DB 15; Length 261;
Best Local Similarity 100.0%; Pred. No. 6.9e-123;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGGGCGCTCTGTGAGGG	60
Db	1	GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGGGCGCTCTGTGAGGG	60
QY	61	GCTGAGCCTCGGGAGCGTTCCTTCTGGGGAGAGCGTTCCTTATCAACACCATCCGGCT	120
Db	61	GCTGAGCCTCGGGAGCGTTCCTTCTGGGGAGAGCGTTCCTTATCAACACCATCCGGCT	120
QY	121	CGTGGGGCAGTCGTGCTCCCTCGGCCACCGGGGAGCCGATGGCCCTTGAAGGGAGCGCT	180
Db	121	CGTGGGGCAGTCGTGCTCCCTCGGCCACCGGGGAGCCGATGGCCCTTGAAGGGAGCGCT	180
QY	181	GCCCGCGCTCCCGGAGGAGCTGGCTTCGAGGGGAGGTGGAGTACAACGGGGGCTT	240

Db 181 GCCCGCCCTGCCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT 240
Qy 241 CCACCTGGCCATCGACGTGGA 261
Db 241 CCACCTGGCCATCGACGTGGA 261

RESULT 2

US-10-224-624-2
; Sequence 2, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GliTEN
; FILE REFERENCE: 54704.8059.US90
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 261
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-224-624-2

Query Match 100.0%; Score 261; DB 15; Length 261;
Best Local Similarity 100.0%; Pred. No. 6.9e-123; Indels 0; Gaps 0;
Matches 261; Conservative 0; Mismatches 0;
Qy 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCCCTGCTGGAGGG 60
Db 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCCCTGCTGGAGGG 60
Qy 61 GCTGAGCCTGGGAGCAGTGTCTCTGGGCGAGACGGTGCCTTCATCAAGACATCCGGCT 120
Db 61 GCTGAGCCTGGGAGCAGTGTCTCTGGGCGAGACGGTGCCTTCATCAAGACATCCGGCT 120
Qy 121 CGTGGCCAGTCGTGCGCCCTCGGCCACCGGGAGCCCGATGCGCTTCAAGAGAGCGGCT 180
Db 121 CGTGGCCAGTCGTGCGCCCTCGGCCACCGGGAGCCCGATGCGCTTCAAGAGAGCGGCT 180
Qy 181 GCCCGCCCTGCCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT 240
Db 181 GCCCGCCCTGCCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT 240
Qy 241 CCACCTGGCCATCGACGTGGA 261
Db 241 CCACCTGGCCATCGACGTGGA 261

RESULT 3

US-10-051-769-3
; Sequence 3, Application US/10051769
; Publication No. US20030044811A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randy D.
; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION
; FILE REFERENCE: 268/260 (RMJ-00-37)
; CURRENT APPLICATION NUMBER: US/10/051,769
; CURRENT FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,160
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 873
; TYPE: DNA
; ORGANISM: Homo Sapiens

US-10-051-769-3

Query Match 100.0%; Score 261; DB 15; Length 873;
Best Local Similarity 100.0%; Pred. No. 5.7e-123; Indels 0; Gaps 0;
Matches 261; Conservative 0; Mismatches 0;
Qy 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCCCTGCTGGAGGG 60
Db 366 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCCCTGCTGGAGGG 425
Qy 61 GCTGAGCCTGGGAGCAGTGTCTCTGGGCGAGACGGTGCCTTCATCAAGACATCCGGCT 120
Db 426 GCTGAGCCTGGGAGCAGTGTCTCTGGGCGAGACGGTGCCTTCATCAAGACATCCGGCT 485
Qy 121 CGTGGCCAGTCGTGCGCCCTCGGCCACCGGGAGCCCGATGCGCTTCAAGAGAGCGGCT 180
Db 486 GCTGAGCCTGGGAGCAGTGTCTCTGGGCGAGACGGTGCCTTCATCAAGACATCCGGCT 545
Qy 181 GCCCGCCCTGCCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT 240
Db 546 GCCCGCCCTGCCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT 605
Qy 241 CCACCTGGCCATCGACGTGGA 261
Db 606 CCACCTGGCCATCGACGTGGA 626

RESULT 4

US-10-224-624-3
; Sequence 3, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: MCKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GliTEN
; FILE REFERENCE: 54704.8059.US90
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 873
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-224-624-3

Query Match 100.0%; Score 261; DB 15; Length 873;
Best Local Similarity 100.0%; Pred. No. 5.7e-123; Indels 0; Gaps 0;
Matches 261; Conservative 0; Mismatches 0;
Qy 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCCCTGCTGGAGGG 60
Db 366 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCCCTGCTGGAGGG 425
Qy 61 GCTGAGCCTGGGAGCAGTGTCTCTGGGCGAGACGGTGCCTTCATCAAGACATCCGGCT 120
Db 426 GCTGAGCCTGGGAGCAGTGTCTCTGGGCGAGACGGTGCCTTCATCAAGACATCCGGCT 485
Qy 121 CGTGGCCAGTCGTGCGCCCTCGGCCACCGGGAGCCCGATGCGCTTCAAGAGAGCGGCT 180
Db 486 GCTGAGCCTGGGAGCAGTGTCTCTGGGCGAGACGGTGCCTTCATCAAGACATCCGGCT 545
Qy 181 GCCCGCCCTGCCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT 240
Db 546 GCCCGCCCTGCCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT 605
Qy 241 CCACCTGGCCATCGACGTGGA 261
Db 606 CCACCTGGCCATCGACGTGGA 626

```

RESULT 5
US-10-051-769-1
; Sequence 1, Application US/10051769
; Publication No. US2003004481A1
; GENERAL INFORMATION:
; APPLICANT: McKINNON, Randy D.
; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION
; FILE REFERENCE: 268/260 (RWJ-00-37)
; CURRENT APPLICATION NUMBER: US/10/051,769
; PRIOR FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,160
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-051-769-1

Query Match      100.0%; Score 261; DB 15; Length 960;
Best Local Similarity 100.0%; Pred. No. 5.6e-123;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GATCAAGGTGGAGTTCGAGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 60
Db      1 GATCAAGGTGGAGTTCGAGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 60
Db      1 GATCAAGGTGGAGTTCGAGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 455
QY      61 GCTGAGCCTGCGGACGTGTTCTTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
Db      61 GCTGAGCCTGCGGACGTGTTCTTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
Db      61 GCTGAGCCTGCGGACGTGTTCTTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 515
QY      121 CGTGGCGCCAGTCTGCGCCCTCGGCCACCGGGGAGCGGATGGCCCTGAAGGGGAGCGCT 180
Db      121 CGTGGCGCCAGTCTGCGCCCTCGGCCACCGGGGAGCGGATGGCCCTGAAGGGGAGCGCT 180
Db      121 CGTGGCGCCAGTCTGCGCCCTCGGCCACCGGGGAGCGGATGGCCCTGAAGGGGAGCGCT 575
QY      181 GCCCGCGCGCTGCGCCCGAGGAGCTGCGCTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 240
Db      181 GCCCGCGCGCTGCGCCCGAGGAGCTGCGCTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 240
Db      181 GCCCGCGCGCTGCGCCCGAGGAGCTGCGCTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 575
QY      576 GCCCGCGCGCTGCGCCCGAGGAGCTGCGCTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 635
Db      576 GCCCGCGCGCTGCGCCCGAGGAGCTGCGCTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 635
QY      241 CCACCTGGCCATCGACGTGGA 261
Db      241 CCACCTGGCCATCGACGTGGA 261
Db      636 CCACCTGGCCATCGACGTGGA 656

RESULT 6
US-10-051-769-1
; Sequence 1, Application US/10051769
; Publication No. US2003004481A1
; GENERAL INFORMATION:
; APPLICANT: McKINNON, Randy D.
; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION
; FILE REFERENCE: 268/260 (RWJ-00-37)
; CURRENT APPLICATION NUMBER: US/10/051,769
; PRIOR FILING DATE: 2001-10-20
; PRIOR APPLICATION NUMBER: US 60/242,160
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-051-769-1

Query Match      100.0%; Score 261; DB 15; Length 960;
Best Local Similarity 100.0%; Pred. No. 5.6e-123;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GATCAAGGTGGAGTTCGAGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 60
Db      1 GATCAAGGTGGAGTTCGAGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 60
Db      1 GATCAAGGTGGAGTTCGAGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 455
QY      61 GCTGAGCCTGCGGACGTGTTCTTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
Db      61 GCTGAGCCTGCGGACGTGTTCTTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
Db      61 GCTGAGCCTGCGGACGTGTTCTTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 515
QY      121 CGTGGCGCCAGTCTGCGCCCTCGGCCACCGGGGAGCGGATGGCCCTGAAGGGGAGCGCT 180
Db      121 CGTGGCGCCAGTCTGCGCCCTCGGCCACCGGGGAGCGGATGGCCCTGAAGGGGAGCGCT 180
Db      121 CGTGGCGCCAGTCTGCGCCCTCGGCCACCGGGGAGCGGATGGCCCTGAAGGGGAGCGCT 575
QY      181 GCCCGCGCGCTGCGCCCGAGGAGCTGCGCTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 240
Db      181 GCCCGCGCGCTGCGCCCGAGGAGCTGCGCTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 240
Db      181 GCCCGCGCGCTGCGCCCGAGGAGCTGCGCTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 575
QY      576 GCCCGCGCGCTGCGCCCGAGGAGCTGCGCTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 635
Db      576 GCCCGCGCGCTGCGCCCGAGGAGCTGCGCTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 635
QY      241 CCACCTGGCCATCGACGTGGA 261
Db      241 CCACCTGGCCATCGACGTGGA 261
Db      636 CCACCTGGCCATCGACGTGGA 656

RESULT 7
US-10-104-047-684
; Sequence 684, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 684
; LENGTH: 2639
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-684

Query Match      100.0%; Score 261; DB 16; Length 2639;
Best Local Similarity 100.0%; Pred. No. 4.7e-123;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GATCAAGGTGGAGTTCGAGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 60
Db      1 GATCAAGGTGGAGTTCGAGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 60
Db      1 GATCAAGGTGGAGTTCGAGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 638
QY      61 GCTGAGCCTGCGGACGTGTTCTTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
Db      61 GCTGAGCCTGCGGACGTGTTCTTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
Db      61 GCTGAGCCTGCGGACGTGTTCTTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 698
QY      121 CGTGGCGCCAGTCTGCGCCCTCGGCCACCGGGGAGCGGATGGCCCTGAAGGGGAGCGCT 180
Db      121 CGTGGCGCCAGTCTGCGCCCTCGGCCACCGGGGAGCGGATGGCCCTGAAGGGGAGCGCT 180
Db      121 CGTGGCGCCAGTCTGCGCCCTCGGCCACCGGGGAGCGGATGGCCCTGAAGGGGAGCGCT 758
QY      181 GCCCGCGCGCTGCGCCCGAGGAGCTGCGCTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 240
Db      181 GCCCGCGCGCTGCGCCCGAGGAGCTGCGCTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 240
Db      181 GCCCGCGCGCTGCGCCCGAGGAGCTGCGCTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 818
QY      241 CCACCTGGCCATCGACGTGGA 261
Db      241 CCACCTGGCCATCGACGTGGA 261
Db      819 CCACCTGGCCATCGACGTGGA 839

RESULT 8
US-10-224-624-9
; Sequence 9, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: McKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein Gliten
; FILE REFERENCE: 54704.8059.US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-10-224-624-1

Query Match      100.0%; Score 261; DB 15; Length 960;
Best Local Similarity 100.0%; Pred. No. 5.6e-123;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GATCAAGGTGGAGTTCGAGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 60
Db      1 GATCAAGGTGGAGTTCGAGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGGAGGG 60
```

; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 3465
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-224-624-9

Query Match 100.0%; Score 261; DB 15; Length 3465;
Best Local Similarity 100.0%; Pred. No. 4.5e-123;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGGTGAGTTTCGAGGAGCTGCTGCAGACCAAGACGCGCGGCGCTGCTGGAGGG 60
DB 366 GATCAAGGTGAGTTTCGAGGAGCTGCTGCAGACCAAGACGCGCGGCGCTGCTGGAGGG 425

QY 61 GCTGAGCTCGGAGACGTGTTCTTGGGCGAGACGCGTGCCTTCATCAAGACCATCCGGCT 120
DB 426 GCTGAGCTCGGAGACGTGTTCTTGGGCGAGACGCGTGCCTTCATCAAGACCATCCGGCT 485

QY 121 CGTGGGCGAGTGTGCTCCCTCGGCGACCGGGAGGCCGATGGCCCTGAAGGGAGGCGCT 180
DB 486 CGTGGGCGAGTGTGCTCCCTCGGCGACCGGGAGGCCGATGGCCCTGAAGGGAGGCGCT 545

QY 181 GCCCGCCCTGCCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT 240
DB 546 GCCCGCCCTGCCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT 605

QY 241 CCACCTGGCCATCGACGTGGA 261
DB 606 CCACCTGGCCATCGACGTGGA 626

RESULT 9
US-10-112-944-63
; Sequence 63, Application US/10112944
; Publication No. US20040048249A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yang, Yonghong
; APPLICANT: Weng, Gezhi
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Xue, Aidong J.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wehrman, Tom
; APPLICANT: Ghosh, Malabika
; APPLICANT: Wang, Dunrui
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Zhiwei
; TITLE OF INVENTION: No. US20040048249A1el Nucleic Acids and
; FILE REFERENCE: 805A
; CURRENT APPLICATION NUMBER: US/10/112,944
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: US 09/519,705
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/540,217
; PRIOR FILING DATE: 2000-03-31

; PRIOR APPLICATION NUMBER: US 09/552,929
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/577,408
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 924
; SOFTWARE: pt_FL_genes Version 5.0
; SEQ ID NO 63
; LENGTH: 3649
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3462)
US-10-112-944-63

Query Match 100.0%; Score 261; DB 13; Length 3649;
Best Local Similarity 100.0%; Pred. No. 4.4e-123;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGGTGAGTTTCGAGGAGCTGCTGCAGACCAAGACGCGCGGCGCTGCTGGAGGG 60
DB 366 GATCAAGGTGAGTTTCGAGGAGCTGCTGCAGACCAAGACGCGCGGCGCTGCTGGAGGG 425

QY 61 GCTGAGCTCGGAGACGTGTTCTTGGGCGAGACGCGTGCCTTCATCAAGACCATCCGGCT 120
DB 426 GCTGAGCTCGGAGACGTGTTCTTGGGCGAGACGCGTGCCTTCATCAAGACCATCCGGCT 485

QY 121 CGTGGGCGAGTGTGCTCCCTCGGCGACCGGGAGGCCGATGGCCCTGAAGGGAGGCGCT 180
DB 486 CGTGGGCGAGTGTGCTCCCTCGGCGACCGGGAGGCCGATGGCCCTGAAGGGAGGCGCT 545

QY 181 GCCCGCCCTGCCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT 240
DB 546 GCCCGCCCTGCCCCGAGGAGCTGGCCCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT 605

QY 241 CCACCTGGCCATCGACGTGGA 261
DB 606 CCACCTGGCCATCGACGTGGA 626

RESULT 10
US-10-224-624-7
; Sequence 7, Application US/10224624
; Publication No. US20030108915A1
; GENERAL INFORMATION:
; APPLICANT: McKINNON, Randall D.
; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GliTEN
; FILE REFERENCE: 54704.8059.US00
; CURRENT APPLICATION NUMBER: US/10/224,624
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 60/242,160
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 10/051,769
; PRIOR FILING DATE: 2001-10-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 3832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (178)..(3639)
; OTHER INFORMATION:
US-10-224-624-7

Query Match 100.0%; Score 261; DB 15; Length 3832;
Best Local Similarity 100.0%; Pred. No. 4.4e-123;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGGTGAGTTTCGAGGAGCTGCTGCAGACCAAGACGCGCGGCGCTGCTGGAGGG 60
DB 543 GATCAAGGTGAGTTTCGAGGAGCTGCTGCAGACCAAGACGCGCGGCGCTGCTGGAGGG 602


```
; FILE REFERENCE: 21402-533A
; CURRENT APPLICATION NUMBER: US/10/336,603A
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: 09/746,491
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 10/055,569
; PRIOR FILING DATE: 2001-10-26
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 25
; LENGTH: 4801
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (178)..(3639)
US-10-336-603A-25
```

```
Query Match 100.0%; Score 261; DB 13; Length 4470;
Best Local Similarity 100.0%; Pred. No. 4.3e-123;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCGCGCTTCAATCAAGACCATCCGGCT 120
Db 366 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCGCGCTTCAATCAAGACCATCCGGCT 425

QY 61 GCTGAGCTCGGGACGTGTTCTTGGCGAGAGCGTGCCCTTCAATCAAGACCATCCGGCT 120
Db 426 GCTGAGCTCGGGACGTGTTCTTGGCGAGAGCGTGCCCTTCAATCAAGACCATCCGGCT 485

QY 121 CGTGGGCCAGTCTGCTCCCTCGGCCACCGGGAGCCGATGGCCCTTCAAGGGAGGCGCT 180
Db 486 CGTGGGCCAGTCTGCTCCCTCGGCCACCGGGAGCCGATGGCCCTTCAAGGGAGGCGCT 545

QY 181 GCCCGCGCTCGCCCGGAGGAGTGGCTTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 240
Db 546 GCCCGCGCTCGCCCGGAGGAGTGGCTTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 605

QY 241 CCACCTGGCCATCGACGTGA 261
Db 606 CCACCTGGCCATCGACGTGA 626

US-10-276-774-950
```

```
Query Match 100.0%; Score 261; DB 13; Length 4470;
Best Local Similarity 100.0%; Pred. No. 4.3e-123;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCGCGCTTCAATCAAGACCATCCGGCT 60
Db 366 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCGCGCTTCAATCAAGACCATCCGGCT 425

QY 61 GCTGAGCTCGGGACGTGTTCTTGGCGAGAGCGTGCCCTTCAATCAAGACCATCCGGCT 120
Db 426 GCTGAGCTCGGGACGTGTTCTTGGCGAGAGCGTGCCCTTCAATCAAGACCATCCGGCT 485

QY 121 CGTGGGCCAGTCTGCTCCCTCGGCCACCGGGAGCCGATGGCCCTTCAAGGGAGGCGCT 180
Db 486 CGTGGGCCAGTCTGCTCCCTCGGCCACCGGGAGCCGATGGCCCTTCAAGGGAGGCGCT 545

QY 181 GCCCGCGCTCGCCCGGAGGAGTGGCTTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 240
Db 546 GCCCGCGCTCGCCCGGAGGAGTGGCTTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 605

QY 241 CCACCTGGCCATCGACGTGA 261
Db 606 CCACCTGGCCATCGACGTGA 626

US-10-276-774-950
```

```
Query Match 100.0%; Score 261; DB 12; Length 4801;
Best Local Similarity 100.0%; Pred. No. 4.2e-123;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCGCGCTTCAATCAAGACCATCCGGCT 60
Db 543 GATCAAGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGCGCGCGCTTCAATCAAGACCATCCGGCT 602

QY 61 GCTGAGCTCGGGACGTGTTCTTGGCGAGAGCGTGCCCTTCAATCAAGACCATCCGGCT 120
Db 603 GCTGAGCTCGGGACGTGTTCTTGGCGAGAGCGTGCCCTTCAATCAAGACCATCCGGCT 662

QY 121 CGTGGGCCAGTCTGCTCCCTCGGCCACCGGGAGCCGATGGCCCTTCAAGGGAGGCGCT 180
Db 663 CGTGGGCCAGTCTGCTCCCTCGGCCACCGGGAGCCGATGGCCCTTCAAGGGAGGCGCT 722

QY 181 GCCCGCGCTCGCCCGGAGGAGTGGCTTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 240
Db 723 GCCCGCGCTCGCCCGGAGGAGTGGCTTTCGAGGCGGAGGTGAGTACAAACGGGGGCTT 782

QY 241 CCACCTGGCCATCGACGTGA 261
Db 783 CCACCTGGCCATCGACGTGA 803

US-10-250-823-11
```

```
Query Match 18.0%; Score 47; DB 12; Length 3847;
Best Local Similarity 100.0%; Pred. No. 3.7e-14;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 GCGGAGGTGGAGTACAAACGGGGGCTTCCACCTGGCCATCGACGTGA 261
Db 1 GCGGAGGTGGAGTACAAACGGGGGCTTCCACCTGGCCATCGACGTGA 47

US-10-250-823-11
```

RESULT 14

US-10-051-769-4
 ; Sequence 4, Application US/10051769
 ; Publication No. US2003004811A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MCKINNON, Randy D.
 ; TITLE OF INVENTION: AN EST-DEFINED PROBE FOR CANCER PROGRESSION
 ; FILE REFERENCE: 268/260 (RWJ-00-37)
 ; CURRENT APPLICATION NUMBER: US/10/051,769
 ; CURRENT FILING DATE: 2001-10-20
 ; PRIOR APPLICATION NUMBER: US 60/242,160
 ; PRIOR FILING DATE: 2000-10-20
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 4
 ; LENGTH: 263
 ; TYPE: DNA
 ; ORGANISM: Sprague Dawley rat
 US-10-051-769-4

Query Match 16.5%; Score 43; DB 15; Length 263;
 Best Local Similarity 100.0%; Pred. No. 6.3e-12;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCC 43
 ||||||||||||||||||||||||||||||||||||||||||||
 Db 1 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCC 43

RESULT 15

US-10-224-624-4
 ; Sequence 4, Application US/10224624
 ; Publication No. US20030108915A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MCKINNON, Randall D.
 ; TITLE OF INVENTION: Glioblastoma Multiforme Associated Protein GLITEN
 ; FILE REFERENCE: 54704.8059.US00
 ; CURRENT APPLICATION NUMBER: US/10/224,624
 ; CURRENT FILING DATE: 2002-08-20
 ; PRIOR APPLICATION NUMBER: 60/242,160
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 10/051,769
 ; PRIOR FILING DATE: 2001-10-20
 ; NUMBER OF SEQ ID NOS: 9
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 4
 ; LENGTH: 263
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-224-624-4

Query Match 16.5%; Score 43; DB 15; Length 263;
 Best Local Similarity 100.0%; Pred. No. 6.3e-12;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCC 43
 ||||||||||||||||||||||||||||||||||||||||||||
 Db 1 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCC 43

Search completed: July 18, 2004, 15:53:46
 Job time : 265 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2004, 11:40:25 ; Search time 2467.95 Seconds
(without alignments)
4583.781 Million cell updates/sec

Title: US-10-051-769-2
Perfect score: 261
Sequence: 1 gatcaagtggtgaggg.....cacotggccatcgacgtgga 261

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_nam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	261	100.0	2623	9	BC012186	BC012186 Homo sapi
2	261	100.0	2639	6	AX747159	AX747159 Sequence
3	261	100.0	2639	9	AK091746	AK091746 Homo sapi
4	261	100.0	3856	9	BC028375	BC028375 Homo sapi
5	261	100.0	49052	9	AL359836	AL359836 Human DNA
6	261	100.0	120578	9	AC005887	AC005887 citb_173
7	227.4	87.1	68196	2	AC108407	AC108407 Mus muscu
8	227.4	87.1	221524	10	AC139040	AC139040 Mus muscu
9	222.6	85.3	243412	2	AC127769	AC127769 Rattus no
10	87.6	33.6	179734	2	BC511030	BC511030 Danio rer
11	71.4	27.4	18666	5	AF469049	AF469049 Gallus ga
12	54.4	20.8	321250	1	SC093911	AL939111 Streptomy
13	51.8	19.8	67702	2	AC134262	AC134262 Rattus no
14	51.8	19.8	219992	2	AC118991	AC118991 Rattus no
15	51.8	19.8	249592	2	AC095429	AC095429 Rattus no
16	51.6	19.8	6436	6	BD271287	BD271287 Polyhydro
17	51.6	19.8	6436	6	AR430598	AR430598 Sequence
18	51.2	19.6	32748	1	AB070951	AB070951 Streptomy
19	51.2	19.6	300800	1	AP005036	AP005036 Streptomy
20	50.8	19.5	241244	2	AC113721	AC113721 Rattus no
21	50	19.2	1812	6	BD180122	BD180122 Highly th
22	49.8	19.1	281450	1	AP005032	AP005032 Streptomy
23	49.8	19.1	300100	1	SC093912	AL939123 Streptomy
24	49.4	18.9	298550	1	AP005047	AP005047 Streptomy
25	49.2	18.9	10029	1	AE005971	AE005971 Caulobact
26	48.8	18.7	205050	1	AL646082	AL646082 Ralstonia
27	48.8	18.7	348257	1	EX640425	EX640425 Bordetell
28	48.8	18.7	349497	1	EX640440	EX640440 Bordetell
29	48.6	18.6	201050	1	AL646064	AL646064 Ralstonia
30	48.4	18.5	136822	5	AC091091	AC091091 Gallus ga
31	48	18.4	1660	8	AK069852	AK069852 Oryza sat
32	48	18.4	157309	2	AP005004	AP005004 Oryza sat
33	48	18.4	309135	1	AP005946	AP005946 Bradyrhiz
34	47.8	18.3	1085	9	BC035849	BC035849 Homo sapi
35	47.8	18.3	2710	6	AX540750	AX540750 Sequence
36	47.8	18.3	2778	9	AK128815	AK128815 Homo sapi
37	47.8	18.3	12248	1	AE001929	AE001929 Deinococc
38	47.8	18.3	47090	9	AC092310	AC092310 Homo sapi
39	47.8	18.3	47090	9	AC093233	AC093233 Homo sapi
40	47.8	18.3	138203	1	AY310323	AY310323 Streptomy
41	47.8	18.3	156942	2	AC136470	AC136470 Homo sapi
42	47.8	18.3	166618	9	AC010323	AC010323 Homo sapi
43	47.8	18.3	313800	1	SC093914	AL939114 Streptomy
44	47.6	18.2	2466	6	BD180170	BD180170 Highly th
45	47.6	18.2	348642	1	EX640446	EX640446 Bordetell

ALIGNMENTS

RESULT 1	BC012186	2623 bp	mRNA	linear	PRI 06-AUG-2001
LOCUS	Homo sapiens				
DEFINITION	Homo sapiens				
ACCESSION	BC012186				
VERSION	BC012186.1	GI:15082555			
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	1 (bases 1 to 2623)				
JOURNAL	Strausberg, R.				
	Direct Submission				
	Submitted (02-AUG-2001) National Institutes of Health, Mammalian				

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabbs-remail.nih.gov

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Rubin Laboratory

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

cDNA Sequencing by: Genome Sequence Centre,

BC Cancer Agency, Vancouver, BC, Canada

info@cgsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 29 Row: h Column: 17
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Genomescan gene prediction.

Location/Qualifiers

1..2623

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4564853"

/tissue_type="Kidney, renal cell adenocarcinoma"

/clone_lib="NIH_MGC_14"

/lab_host="DH10B-R"

/note="Vector: pOTB7"

ORIGIN

Query Match 100.0%; Score 261; DB 9; Length 2623;
Best Local Similarity 100.0%; Pred. No. 2.2e-31;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGTGGAGTTTCGAGGAGCTCTGCAGACCAAGACGGCGCGGCGCTGCTGGAGGG 60
DB 543 GATCAAGTGGAGTTTCGAGGAGCTCTGCAGACCAAGACGGCGCGGCGCTGCTGGAGGG 602
QY 61 GCTGAGCTCGGGACAGTGTTCCTTGGCGGACAGCGTGCCTTCATCAAGACCATCCGGCT 120
DB 603 GCTGAGCTCGGGACAGTGTTCCTTGGCGGACAGCGTGCCTTCATCAAGACCATCCGGCT 662
QY 121 CGTGGGCGAGTCTGGCCCTCGCCACCGGAGGCGCGATGCCCTGAAGGGAGGCGCT 180
DB 663 CGTGGGCGAGTCTGGCCCTCGCCACCGGAGGCGCGATGCCCTGAAGGGAGGCGCT 722
QY 181 GCCCGCGCGCTGCCCGGAGGAGTGGCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT 240
DB 723 GCCCGCGCGCTGCCCGGAGGAGTGGCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT 782
QY 241 CCACCTGGCCATCGACGTGGA 261
DB 783 CCACCTGGCCATCGACGTGGA 803

RESULT 2
AX747159
LOCUS AX747159 2639 bp mRNA linear PAT 20-JUN-2003
DEFINITION Sequence 684 from Patent EP1308459.
ACCESSION AX747159
VERSION AX747159.1 GI:32131547
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S., Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R., Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and Masuho,Y.

Full-length cDNA sequences

Patent: EP 1308459-A 684 07-MAY-2003;

Helix Research Institute (JP) ; Research Association for

Biotechnology (JP)

FEATURES

source

1..2639
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 261; DB 6; Length 2639;
Best Local Similarity 100.0%; Pred. No. 2.2e-31;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCAAGTGGAGTTTCGAGGAGCTCTGCAGACCAAGACGGCGCGGCGCTGCTGGAGGG 60
DB 579 GATCAAGTGGAGTTTCGAGGAGCTCTGCAGACCAAGACGGCGCGGCGCTGCTGGAGGG 638
QY 61 GCTGAGCTCGGGACAGTGTTCCTTGGCGGACAGCGTGCCTTCATCAAGACCATCCGGCT 120
DB 639 GCTGAGCTCGGGACAGTGTTCCTTGGCGGACAGCGTGCCTTCATCAAGACCATCCGGCT 698
QY 121 CGTGGGCGAGTCTGGCCCTCGCCACCGGAGGCGCGATGCCCTGAAGGGAGGCGCT 180
DB 699 CGTGGGCGAGTCTGGCCCTCGCCACCGGAGGCGCGATGCCCTGAAGGGAGGCGCT 758
QY 181 GCCCGCGCGCTGCCCGGAGGAGTGGCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT 240
DB 759 GCCCGCGCGCTGCCCGGAGGAGTGGCTTCGAGCGGAGGTGGAGTACAAACGGGGGCTT 818
QY 241 CCACCTGGCCATCGACGTGGA 261
DB 819 CCACCTGGCCATCGACGTGGA 839

RESULT 3
AX747146
LOCUS Homo sapiens cDNA FLJ34427 fis, clone HHDPC200816. PRI 15-JUL-2002
DEFINITION AK091746
ACCESSION AK091746
VERSION AK091746.1 GI:21750192
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 Nishi,T., Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.

Unpublished

NEDO human cDNA sequencing project

TITLE

JOURNAL

REFERENCE

2 (bases 1 to 2639)

Isogai,T. and Yamamoto,J.

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library

construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5' - & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES
source
1..2639
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HHDPC2008816"
/cell_type="dermal papilla cells (HDPIC)"
/clone_lib="HHDPC2"
/note="Cloning vector: pME18SFL3-primary culture, dermal papilla cells"
ORIGIN
Query Match 100.0%; Score 261; DB 9; Length 2639;
Best Local Similarity 100.0%; Pred. No. 2.2e-31;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCAAGGTGGAGTTCGAGGAGCTGTCGACAGCAAGACGGCGCGCTGCTGGAGGG 60
|||||
Db 579 GATCAAGGTGGAGTTCGAGGAGCTGTCGACAGCAAGACGGCGCGCTGCTGGAGGG 638
QY 61 GCTGAGCTCGGGAGCGTGTCTCGGGCGAGCGGTCCCTTCATCAAGACCATCGGCT 120
|||||
Db 639 GCTGAGCTCGGGAGCGTGTCTCGGGCGAGCGGTCCCTTCATCAAGACCATCGGCT 698
QY 121 CGTGGCGGACGTCGTGCTCGGCCCGCGGAGCGCGATGCGCGTGAAGGGGAGCGCT 180
|||||
Db 699 CGTGGCGGACGTCGTGCTCGGCCCGCGGAGCGCGATGCGCGTGAAGGGGAGCGCT 758
QY 181 GCCCGCGCGCTGCCCGGAGGAGCTGCGCTTCGAGGCGGAGGTGAGTACAAACGGGGCTT 240
|||||
Db 759 GCCCGCGCGCTGCCCGGAGGAGCTGCGCTTCGAGGCGGAGGTGAGTACAAACGGGGCTT 818
QY 241 CCACCTGGCGCATCGAGTGA 261
Db 819 CCACCTGGCGCATCGAGTGA 839

RESULT 4
BC028375
LOCUS
DEFINITION
Homo sapiens hypothetical protein LOC118987, mRNA (cDNA clone MGC:27107 IMAGE:4837939), complete cds.
ACCESSION
BC028375
VERSION
BC028375.1 GI:22382223
KEYWORDS
MGC.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 3856)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raja,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S., Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.C., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smallos,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

TITLE
JOURNAL
MEDLINE
PUBMED
12477932
REFERENCE
2 (bases 1 to 3856)
Strausberg,R.
AUTHORS
DIRECT SUBMISSION
TITLE
JOURNAL
REMARK
COMMENT
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 3856)
Strausberg,R.
Direct Submission
Submitted (23-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 34 Row: m Column: 21
This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction.
Location/Qualifiers
1..3856
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:27107 IMAGE:4837939"
/tissue_type="testis"
/clone_lib="NIH MGC 97"
/lab_host="DH10B"
/notes="Vector: pBluescript"
1..3856
/gene="LOC118987"
/db_xref="LocusID:118987"
203..3667
/codon_start=1
/product="hypothetical protein LOC118987"
/protein_id="AAH28375.1"
/db_xref="GI:22382224"
/db_xref="LocusID:118987"
/translation="MGLLLMLASAVLGSFLTLAQLFLLYRQPEPPADEAARAGEG
FYIKVPGLLLREYLYGGDRDEPSGAAPGGATTAAPETPAPPTRETCYFLNATI
LFLFRDLRTALTTRRWTKKIKVFEPELLQTKTAGRLGLSLRDLVFLGTVFPIKTI
RLVRPVPSATGEPDGEALPAACEELAFEAVEYNGHLAIDVLVFKSAVL
FKLSVVGELRVFTVPFTWFFSFVEDPLIDFEVRSQFEGEPQWLTSLVNLK
KIKRKHUPNYKIRFPFPFTQLOGFEDEHHIQQWALTEGRLKVTLLCSRL
IFGSDREANVHCTLELSSVMEKRSIKTVLEIKGNLSVGLRLRLVSDYDYGAG
HVIETVAPAAIADQGRDLIAIGVKITSTLQVLLIKLQAGDRLVLYVYRPG
QNGQVAFQNFQLEENFSLSCSCSGEYEAAGLTVDTESRELDSEFEDLASDRAQ
NFKDBAQLSHSKPRVPTLSIKPLGAI SPVLRKLVAGSHPLPKIKQKGNKPP
LKTSETDPAQVSKPTQGSFAKVPVPRPOAKVPLPSADAPNAQEPDLVEKPKVP
PVLVDSAEKQANVDAIDAAAAPQFLAKQEVAKDVTSETSCPTKSDSDROTWESS
EILYRNKLGKTRFRASCLFDIACHRYNLIALWCKDPFKGLICLGHVSUGLEDA
LGCLATNTEYLSKRLFAEPKAIIVTRALNLSMQGKDFNKGFCYGDITTHFKYLKE
GSDSHVVTNVEKEPHELVVEVSLPKEEFQVGMGLTENKHSFOQTQFRTWCDY
CKKHVTKAASOCMFCAYCHKKCOBKCLAETSVCAGTDRRIDRTLNLRLEQETLL
GLPPVDAEASKVNTKTLGTRHINTSSRLNLRVSKTRISEPCTDLVPSKPTP
NTSDNGSTEVCGNPSKRGNTGILKVRKEGGLDDSVFAVKEIGRLIRGLPTE
ERIKLEFMDLQNEIDQELHNNLSVREKETTDTTRKSLLSAALASGERLQALT
LMHYRAGVIEDIETLESLSLQHSKKISKYTDOTDEEDLNEISLQIDSPFSSISDD
LFGPSSSV"
1292..1546
/note="PDZ; Region: Domain present in PSD-95, Dlg, and

20-1/2. Also called DHR (Dlg homologous region) or GLGF (relatively well conserved tetrapeptide in these domains). Some PDZs have been shown to bind C-terminal polypeptides" /db_xref="CDD:smart00228"

misc feature

2723..2851
/note="DAG-PE-Bind; Region: Phorbol esters/diacylglycerol binding domain (C1 domain). This domain is also known as the Protein kinase C conserved region 1 (C1) domain" /db_xref="CDD:pfam00130"

ORIGIN

Query Match 100.0%; Score 261; DB 9; Length 3856;
Best Local Similarity 100.0%; Pred. No. 2e-31; Indels 0; Gaps 0;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCAAGTGGAGTTTCGAGGAGTCTGTCAGACCAAGACGGCGCGCTCTCTGAGGG 60
Db 568 GATCAAGTGGAGTTTCGAGGAGTCTGTCAGACCAAGACGGCGCGCTCTCTGAGGG 627
QY 61 GCTGAGCTCGGGACGTCTCTGCGGAGACGCTGCTTCATCAAGACCATCCGGCT 120
Db 628 GCTGAGCTCGGGACGTCTCTGCGGAGACGCTGCTTCATCAAGACCATCCGGCT 687
QY 121 CTTGCGGCGCAGTCGTCTGCGCCACCGGGAGCGCGATGCGCTCAAGGGAGCGCT 180
Db 688 CTTGCGGCGCAGTCGTCTGCGCCACCGGGAGCGCGATGCGCTCAAGGGAGCGCT 747
QY 181 GCCCGCGCTCTGCGGAGAGTCTGCTGAGCGGAGGTGGAGTACAAACGGGGCTT 240
Db 748 GCCCGCGCTCTGCGGAGAGTCTGCTGAGCGGAGGTGGAGTACAAACGGGGCTT 807
QY 241 CCACCTGGCCATCGACGTGGA 261
Db 808 CCACCTGGCCATCGACGTGGA 828

RESULT 5

AL359836/c
LOCUS Human DNA sequence from clone RP11-389E6 on chromosome 10, complete sequence.
DEFINITION AL359836
ACCESSION AL359836
VERSION AL359836.16 GI:1797720
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Smith, M.
Direct Submission
Submitted (21-DEC-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Dec 23, 2001 this sequence version replaced gi:17384082.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPep; Information on the WormPep database can be found at
http://www.sanger.ac.uk/Projects/c_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr10>
RP11-389E6 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-389E6 it may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true left end of clone CTA-109p11 is at 47053 in this sequence. The true right end of clone RP11-129M16 is at 2000 in this sequence.

FEATURES

Location/Qualifiers
1..49052
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-389E6"
/clone_lib="RPCI-11.2"

ORIGIN

Query Match 100.0%; Score 261; DB 9; Length 49052;
Best Local Similarity 100.0%; Pred. No. 1.1e-31; Mismatches 0; Indels 0; Gaps 0;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATCAAGTGGAGTTTCGAGGAGTCTGTCAGACCAAGACGGCGCGCTCTCTGAGGG 60
Db 29767 GATCAAGTGGAGTTTCGAGGAGTCTGTCAGACCAAGACGGCGCGCTCTCTGAGGG 29708
QY 61 GCTGAGCTCGGGACGTCTCTGCGGAGACGCTGCTTCATCAAGACCATCCGGCT 120
Db 29707 GCTGAGCTCGGGACGTCTCTGCGGAGACGCTGCTTCATCAAGACCATCCGGCT 29648
QY 121 CTTGCGGCGCAGTCGTCTGCGCCACCGGGAGCGCGATGCGCTCAAGGGAGCGCT 180
Db 29647 CTTGCGGCGCAGTCGTCTGCGCCACCGGGAGCGCGATGCGCTCAAGGGAGCGCT 29588
QY 181 GCCCGCGCTCTGCGGAGAGTCTGCTGAGCGGAGGTGGAGTACAAACGGGGCTT 240
Db 29587 GCCCGCGCTCTGCGGAGAGTCTGCTGAGCGGAGGTGGAGTACAAACGGGGCTT 29528
QY 241 CCACCTGGCCATCGACGTGGA 261
Db 29527 CCACCTGGCCATCGACGTGGA 29507

RESULT 6

AC005887
LOCUS Human DNA sequence from clone RP11-129M16 on chromosome 10, complete sequence.
DEFINITION AC005887
ACCESSION AC005887
VERSION AC005887.3 GI:6249675
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Smith, D.R.
Sequencing of Human Chromosome 10
Unpublished
2 (bases 1 to 120578)
AUTHORS Smith, D.R.
TITLE Direct Submission
JOURNAL Submitted (29-OCT-1998) Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154, USA
REFERENCE 3 (bases 1 to 120578)
AUTHORS Smith, D.R.
TITLE Direct Submission
JOURNAL Submitted (11-DEC-1998) Genome Therapeutics Corporation, 100 Beaver

```
REFERENCE
AUTHORS      Street, Waltham, MA 02154, USA
TITLE        4 (bases 1 to 120578)
JOURNAL      Direct Submission
JOURNAL      Submitted (02-MAR-1999) Genome Therapeutics Corporation, 100 Beaver
REFERENCE    Street, Waltham, MA 02154, USA
AUTHORS      5 (bases 1 to 120578)
TITLE        Smith,D.R.
JOURNAL      Direct Submission
JOURNAL      Submitted (05-NOV-1999) Genome Therapeutics Corporation, 100 Beaver
REMARK      Street, Waltham, MA 02154, USA
COMMENT      Vector Sequence Clipped
FEATURES     On Nov 5, 1999 this sequence version replaced gi:4314331.
source       Location/Qualifiers
            1..120578
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="10"
            /map="10q25"
            /clone_lib="CIT9878K-1173112"

ORIGIN
Query Match      100.0%; Score 261; DB 9; Length 120578;
Best Local Similarity 100.0%; Pred. No. 8.6e-32;
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCAAGTGGAGTTCGAGGAGTGTCTGCAGACCAAGACGCCGGCGCTGCTGGAGGG 60
Db 53976 GATCAAGTGGAGTTCGAGGAGTGTCTGCAGACCAAGACGCCGGCGCTGCTGGAGGG 54035

QY 61 GCTGAGCTTCGGGAGCGTGTCTTCTGGGCGAGAGCGTGCCTTTCATCAAGACCATCCGGCT 120
Db 54036 GCTGAGCTTCGGGAGCGTGTCTTCTGGGCGAGAGCGTGCCTTTCATCAAGACCATCCGGCT 54095

QY 121 CGTGGCGGACGTCTGCTCCCTCGGCCACCGGGGAGCCGATGGCCCTCAAGGGGAGCGGCT 180
Db 54096 CGTGGCGGACGTCTGCTCCCTCGGCCACCGGGGAGCCGATGGCCCTCAAGGGGAGCGGCT 54155

QY 181 GCCCGCGCTGCCCCGAGGAGTGGCTTCGAGGCGGAGGTGAGTACACGGGGGCTT 240
Db 54156 GCCCGCGCTGCCCCGAGGAGTGGCTTCGAGGCGGAGGTGAGTACACGGGGGCTT 54215

QY 241 CCACCTGGCCATCGACGTGA 261
Db 54216 CCACCTGGCCATCGACGTGA 54236

RESULT 7
AC108407/c
LOCUS      AC108407      68196 bp      DNA      linear      HTG 27-JAN-2002
DEFINITION Mus musculus clone RP24-422P10, LOW-PASS SEQUENCE SAMPLING.
ACCESSION  AC108407
VERSION     AC108407.1 GI:18377216
KEYWORDS   HTG; HTGS PHASE0.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 68196)
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE     Mus musculus, clone RP24-422P10
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 68196)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R.,
```

```
Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zairoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (27-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L18881
Center clone name: 422_P10
-----
* NOTE: This record contains 85 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 714: contig of 714 bp in length
* 715 814: gap of 100 bp
* 815 1520: contig of 706 bp in length
* 1521 1620: gap of 100 bp
* 1621 2323: contig of 703 bp in length
* 2324 2424: gap of 100 bp
* 2424 3134: contig of 711 bp in length
* 3135 3234: gap of 100 bp
* 3235 3937: contig of 703 bp in length
* 3938 4037: gap of 100 bp
* 4038 4740: contig of 702 bp in length
* 4740 4839: gap of 100 bp
* 4839 5545: contig of 706 bp in length
* 5546 5646: gap of 100 bp
* 5646 6352: contig of 707 bp in length
* 6353 6452: gap of 100 bp
* 6453 7172: contig of 720 bp in length
* 7173 7272: gap of 100 bp
* 7273 7996: contig of 724 bp in length
* 7997 8096: gap of 100 bp
* 8097 8814: contig of 718 bp in length
* 8815 9629: contig of 715 bp in length
* 9630 9729: gap of 100 bp
* 9730 10417: contig of 688 bp in length
* 10418 10517: gap of 100 bp
* 10518 11223: contig of 706 bp in length
* 11224 11323: gap of 100 bp
* 11324 12014: contig of 691 bp in length
* 12015 12114: gap of 100 bp
* 12115 12816: contig of 702 bp in length
* 12817 12916: gap of 100 bp
* 12917 13616: contig of 700 bp in length
* 13617 13716: gap of 100 bp
```

```
* 13717 14422: contig of 706 bp in length
* 14423 14522: gap of 100 bp
* 14523 15234: contig of 712 bp in length
* 15234 15334: gap of 100 bp
* 15335 16043: contig of 709 bp in length
* 16044 16143: gap of 100 bp
* 16144 16859: contig of 716 bp in length
* 16860 16969: gap of 100 bp
* 16969 17669: contig of 710 bp in length
* 17670 17769: gap of 100 bp
* 17770 18460: contig of 691 bp in length
* 18461 18560: gap of 100 bp
* 18561 19269: contig of 709 bp in length
* 19270 19369: gap of 100 bp
* 19370 20070: contig of 701 bp in length
* 20071 20170: gap of 100 bp
* 20171 20874: contig of 704 bp in length
* 20875 20974: gap of 100 bp
* 20975 21684: contig of 710 bp in length
* 21685 21784: gap of 100 bp
* 21785 22491: contig of 707 bp in length
* 22492 22591: gap of 100 bp
* 22592 23299: contig of 708 bp in length
* 23300 23399: gap of 100 bp
* 23400 24086: contig of 687 bp in length
* 24087 24186: gap of 100 bp
* 24187 24901: contig of 715 bp in length
* 24902 25001: gap of 100 bp
* 25002 25714: contig of 713 bp in length
* 25715 25814: gap of 100 bp
* 25815 26528: contig of 714 bp in length
* 26529 26628: gap of 100 bp
* 26629 27358: contig of 730 bp in length
* 27359 27458: gap of 100 bp
* 27459 28159: contig of 701 bp in length
* 28160 28259: gap of 100 bp
* 28260 28966: contig of 707 bp in length
* 28967 29066: gap of 100 bp
* 29067 29765: contig of 699 bp in length
* 29766 29865: gap of 100 bp
* 29866 30577: contig of 712 bp in length
* 30578 30677: gap of 100 bp
* 30678 31381: contig of 704 bp in length
* 31382 31481: gap of 100 bp
* 31482 32192: contig of 711 bp in length
* 32193 32292: gap of 100 bp
* 32293 32974: contig of 682 bp in length
* 32975 33074: gap of 100 bp
* 33075 33760: contig of 686 bp in length
* 33761 33860: gap of 100 bp
* 33861 34574: contig of 714 bp in length
* 34575 34674: gap of 100 bp
* 34675 35378: contig of 704 bp in length
* 35379 35478: gap of 100 bp
* 35479 36194: contig of 716 bp in length
* 36195 36294: gap of 100 bp
* 36295 36973: contig of 679 bp in length
* 36974 37073: gap of 100 bp
* 37074 37774: contig of 701 bp in length
* 37775 37874: gap of 100 bp
* 37875 38586: contig of 712 bp in length
* 38587 38686: gap of 100 bp
* 38687 39388: contig of 702 bp in length
* 39389 39488: gap of 100 bp
* 39489 40195: contig of 707 bp in length
* 40196 40295: gap of 100 bp
* 40296 41013: contig of 718 bp in length
* 41014 41113: gap of 100 bp
* 41114 41828: contig of 715 bp in length
* 41829 41928: gap of 100 bp
* 41929 42643: contig of 715 bp in length
* 42644 42743: gap of 100 bp
* 42744 43458: contig of 715 bp in length
```

```
* 43459 43558: gap of 100 bp
* 43559 44242: contig of 684 bp in length
* 44243 44342: gap of 100 bp
* 44343 45043: contig of 701 bp in length
* 45044 45143: gap of 100 bp
* 45144 45856: contig of 713 bp in length
* 45857 45956: gap of 100 bp
* 45957 46654: contig of 698 bp in length
* 46655 46754: gap of 100 bp
* 46755 47449: contig of 695 bp in length
* 47450 47549: gap of 100 bp
* 47550 48258: contig of 709 bp in length
* 48259 48358: gap of 100 bp
* 48359 49070: contig of 712 bp in length
* 49071 49170: gap of 100 bp
* 49171 49886: contig of 716 bp in length
* 49887 49986: gap of 100 bp
* 49987 50673: contig of 687 bp in length
* 50674 50773: gap of 100 bp
* 50774 51472: contig of 699 bp in length
* 51473 51572: gap of 100 bp
* 51573 52288: contig of 716 bp in length
* 52289 52388: gap of 100 bp
* 52389 53075: contig of 687 bp in length
* 53076 53175: gap of 100 bp
* 53176 53879: contig of 704 bp in length
* 53880 53979: gap of 100 bp
* 53980 54675: contig of 696 bp in length
* 54676 54775: gap of 100 bp
* 54776 55477: contig of 702 bp in length
* 55478 55577: gap of 100 bp

Query Match      87.1%; Score 227.4; DB 2; Length 68196;
Best Local Similarity 92.0%; Pred. No. 1.6e-26;
Matches 240; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 GATCAAGGTGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGGCTGCTGGAGGG 60
   |||||
Db 17505 GATCAAGGTGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGGCTGCTGGAGGG 17446

QY 61 GCTGAGCCTGCGGGACGTGTTCTTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 120
   |||||
Db 17445 GCTGAGTCTGCGGACGTGTTCTTGGGCGAGACGGTGCCTTCATCAAGACCATCCGGCT 17386

QY 121 CGTGGCGCACTGCTGCCCTGCGCACCGGGAGCGCGATGCCCTGAAGGGAGCGCGCT 180
   |||||
Db 17385 GGTGCGGCCCTGTGTGGCTCGGGCACCGCGGAGCGCGACCGCGGACCGCGCT 17326

QY 181 GCCCGCCGCTGCCCCGAGGAGCTGCTCGAGCGGAGGTGGAGTACAAAGGGGGCTT 240
   |||||
Db 17325 GCCCGCCACCTGCCCGGAGGAGCTGGCCCTTTGAAGCGGAGGTGGAGTACAAAGGGGGCTT 17266

QY 241 CCACCTGCCCATCGACGTGGA 261
   |||||
Db 17265 CCACCTGCCCATCGACGTGGA 17245

RESULT 8
AC139040
LOCUS      221524 bp DNA linear ROD 16-JUL-2003
DEFINITION Mus musculus chromosome 19, clone RP23-292H20, complete sequence.
ACCESSION AC139040
VERSION   AC139040.7 GI:32813610
KEYWORDS  HTG.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 221524)
AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
TITLE     Mus musculus chromosome 19, clone RP23-292H20
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 221524)
```


AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (23-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 221524)

REFERENCE
AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (31-MAY-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 221524)

REFERENCE
AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (16-JUL-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 16, 2003 this sequence version replaced gi:31249817.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L29063
Center clone name: 292_H_20

FEATURES	Location/Qualifiers
source	1..221524
	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/chromosome="19"
	/map="19"
	/clone="RP23-292H20"
	/clone_lib="RPCI-23 Female Mouse BAC"
	1..7046
	/note="wgs end extension"
	clone end:SP6
	1234..1239
	/note="<30 qual SNCL region"
	complement(1312..1743)
	/rpt_family="MLT1H1"
	complement(1981..2123)
	/rpt_family="MER117"
	2143..2511
	/rpt_family="ORR1D"
	2897..2955
	/rpt_family="(CA)n"
	3063..3114
	/rpt_family="(TG)n"
	3141..3217
	/rpt_family="(CA)n"
	complement(3235..3402)
	/rpt_family="B4A"
	complement(3447..3584)
	/rpt_family="B1 MM"
	complement(3456..3690)
	/rpt_family="ID_B1"
	3724..3752
	/rpt_family="(A)n"
	complement(4335..4351)
	/rpt_family="B3A"
	4352..4408
	/rpt_family="(TG)n"
	complement(4409..4625)
	/rpt_family="B3A"
	complement(4701..4799)
	/rpt_family="WIR"
	complement(5000..5112)
	/rpt_family="B1F"
	5288..5431
	/rpt_family="ORR1D"
	5782..5870
	/rpt_family="B4"
	5897..6040
	/rpt_family="B1_MM"
	6447..6481
	/rpt_family="(TC)n"
	6481..6516
	/rpt_family="(CA)n"
	7047..7052
	/note="clone boundary"
	clone end:SP6
	site:ECORI
	7196..7246
	/rpt_family="(GAAA)n"
	complement(7364..7656)
	/rpt_family="ORR1C"

(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GZSV
Center clone name: CH230-157C16
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 220496 bases at least Q40
Consensus quality: 22898 bases at least Q30
Consensus quality: 224623 bases at least Q20
Estimated insert size: 224491; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 11919: contig of 11919 bp in length
* 11920 12019: gap of unknown length
* 12020 18301: contig of 6282 bp in length
* 18302 18401: gap of unknown length
* 18402 243412: contig of 225011 bp in length.
----- Location/Qualifiers
1. .243412
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-157C16"
1. .1535
/note="wgs contig"
12020. .13436
/note="wgs contig"
17209. .18301
/note="wgs contig"
18402. .13442
/note="wgs contig"
25075. .27269
/note="wgs contig"
238199. .242288
/note="wgs contig"
242339. .243412
/note="wgs contig"

ORIGIN
Query Match 85.3%; Score 222.6; DB 2; Length 243412;
Best Local Similarity 90.8%; Pred. No. 6.3e-26;
Matches 237; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
Qy 1 GATCAAGTGGAGTTTCGAGGAGCTGCTCAGACCAAGACGCCGCGCTGCTGGAGGG 60
Db 90830 GATCAAGTGGAGTTTCGAGGAGCTGCTCAGACCAAGACGCCGCGCTGCTGGAGGG 90889
Qy 61 GCTGAGCTGGCGGACGTGTTCTCGGGGAGACGGTGCCTTCATCAAGACATCCGGCT 120

Db 90890 GCTGAGCTGGCGGACGTGTTCTCGGGTGACACCGTGCCCTACATCAGACCATCCGGCT 90949
Qy 121 COTGCGGCCAGTGTGCTCCCTCGGCCACCGGAGGCCGATGSCCTCGAAGGGGAGCGCT 180
Db 90950 GGTGCGGCCGCTGTGCTTCCGGCCACCGCGACGCCGACGACCGGAGCGCTCT 91009
Qy 181 GCCCGCGCTCCCGGAGGAGCTGSCCTTCGAGGCGGAGGTGGAGTACAAACGGGGGCTT 240
Db 91010 GCCCGCACCTGCCCGGAGGAGCTGSCCTTCGAGGCGGAGGTGGAGTACAAACGGCGGCTT 91069
Qy 241 CCACCTGGCCATCGACGTGGA 261
Db 91070 CCACCTGGCCATCGACGTGGA 91090
RESULT 10
BX511030 179734 bp DNA linear HTG 30-JUN-2003
LOCUS Danio rerio clone CH211-66120, WORKING DRAFT SEQUENCE, 3 unordered
DEFINITION pieces.
ACCESSION BX511030
VERSION BX511030.3 GI:32399530
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 179734)
McLaren, S.
Direct Submission
Submitted (26-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CH10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jul 1, 2003 this sequence version replaced gi:31071341.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zC66120
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 178170 bases at least Q40
Consensus quality: 178389 bases at least Q30
Consensus quality: 178746 bases at least Q20
Insert size: 179534; sum-of-contigs
Insert size: 181688; 4.5% error; agarose-fp
Quality coverage: 9.11x in Q20 bases; sum-of-contigs Quality
coverage: 9.46x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 89039: contig of 89039 bp in length
* 89040 89139: gap of 100 bp
* 89140 169174: contig of 80035 bp in length
* 169175 169274: gap of 100 bp
* 169275 179734: contig of 10460 bp in length.
----- Location/Qualifiers
1. .179734
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-66120"
/clone_lib="CHORI-211"

FEATURES
source

```
misc_feature
1. .89039
/note="assembly fragment:02405
fragment chain:1
clone_end:T7
vector_side:left"
89140..169174
/misc_feature
/note="assembly fragment:01570
fragment chain:1
clone_end:SP6
vector_side:right"
169275..179734
/misc_feature
/note="assembly fragment:00769"
179734..180000

ORIGIN
Query Match 33.6%; Score 87.6; DB 2; Length 179734;
Best Local Similarity 78.4%; Pred. No. 5.1e-05;
Matches 105; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCGCTGCTGCAGGG 60
DB 167240 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCGCTGCTGCAGGG 167299

QY 61 GCTGAGCCTCGCGGAGCTGCTGCGGCGAGACGGTCCCTTCATCAAGACCATCCGGCT 120
DB 167300 GCTGAGCCTCGCGGAGCTGCTGCGGCGAGACGGTCCCTTCATCAAGACCATCCGGCT 167359

QY 121 CGTGGCGGCGAGTCG 134
DB 167360 AATGAACCCGTCG 167373

RESULT 11
AF469049/c
LOCUS
AF469049 18666 bp DNA linear VRT 18-APR-2002
DEFINITION
Gallus gallus caspase-6 gene, complete cds.
ACCESSION
AF469049
VERSION
AF469049.1 GI:18920651
KEYWORDS
Gallus gallus (chicken)
SOURCE
Gallus gallus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE
1 (bases 1 to 18666)
Ruchaud,S., Korfali,N., Villa,P., Kottke,T.J., Dingwall,C.,
Kaufmann,S.H. and Earnshaw,W.C.
Gallus gallus caspase-6 gene disruption reveals a requirement for lamin A
cleavage in apoptotic chromatin condensation
EMBO J. 21 (8), 1967-1977 (2002)
JOURNAL
MEDLINE
21950342
PUBMED
11953316
REFERENCE
2 (bases 1 to 18666)
Ruchaud,S., Korfali,N., Villa,P. and Earnshaw,W.C.
Direct Submission
AUTHORS
Submitted (16-JUN-2002) ICMB, Edinburgh University, Mayfield Road,
Edinburgh EH9 3UR, UK
JOURNAL
FEATURES
Location/Qualifiers
source
1. 18666
/organism="Gallus gallus"
/mol_type="genomic DNA"
/db_xref="taxon:9031"
join(6628..6658,7910..7951,9143..9188,11230..11376,
11816..11892,12336..12511,13158..13314,13938..14176)
/product="caspase-6"
join(6628..6658,7910..7951,9143..9188,11230..11376,
11816..11892,12336..12511,13158..13314,13938..14176)
/codon_start=1
/product="caspase-6"
/protein_id="AAL82386.1"
/db_xref="GI:18920652"
/translation="MSGARRPAAAGRVQLDSKPTPTTADGNQNI TEVDAFDRKRTFD
PAEQYKNHQRGVALIFNHEHFWHLRLPDRGTLADRNKLSLIDLGEVFRIFDD
LKAEVLKVFESRSDYSNADCFVFLSGENDHVIYADQIKIETITNMRGRDC

misc_feature
1. .89039
/note="assembly fragment:02405
fragment chain:1
clone_end:T7
vector_side:left"
89140..169174
/misc_feature
/note="assembly fragment:01570
fragment chain:1
clone_end:SP6
vector_side:right"
169275..179734
/misc_feature
/note="assembly fragment:00769"
179734..180000

ORIGIN
Query Match 27.4%; Score 71.4; DB 5; Length 18666;
Best Local Similarity 65.2%; Pred. No. 0.029;
Matches 105; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 1 GATCAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAAGACGGCGCGCTGCTGCAGGG 60
DB 1361 GATTAAGGTGGAGTTCGAGGAGCTGCTGCAGACCAATTCACAGGAAGGTGCTGGAGAA 1302

QY 61 GCTGAGCCTCGCGGAGCTGCTGCGGCGAGACGGTCCCTTCATCAAGACCATCCGGCT 120
DB 1301 GCTGAGCCTCGCGGAGCTGCTGCGGCGAGACGGTCCCTTCATCAAGACCATCCGGCT 1242

QY 121 CGTGGCGGCGAGTCGCTGCGGCGAGACGGTCCCTTCATCAAGACCATCCGGCT 161
DB 1241 CATCTGGCTGGTCTGCTGCAAGGAGAGAGGTGCCCAAG 1201

RESULT 12
SC0939111
LOCUS
SC0939111 321250 bp DNA linear BCT 11-FEB-2003
DEFINITION
Streptomyces coelicolor A3(2) complete genome; segment 8/29.
ACCESSION
AL939111 AL049497 AL096872 AL109663 AL109663 AL109732
AL109849 AL133219 AL133220 AL136500 AL355752 AL445343
AL590743 AL592262 AL645882
AL939111.1 GI:24419015
VERSION
Streptomyces coelicolor A3(2)
KEYWORDS
Streptomyces coelicolor A3(2)
SOURCE
Streptomyces coelicolor A3(2)
ORGANISM
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
1
Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L.,
Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H.,
Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
Huang,C.H., Kieser,T., Larke,L., Larke,L., Murphy,L., Oliver,K., O'Neill,S.,
Rabinowitch,E., Rajandream,M.A., Rutherford,K., Rutter,S.,
Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S.,
Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrell,B.G.,
Parkhill,J. and Hopwood,D.A.
Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2)
Nature 417 (6885), 141-147 (2002)
JOURNAL
MEDLINE
21996410
PUBMED
12000953
REFERENCE
2 (bases 1 to 321250)
Bentley,S.D.
Direct Submission
AUTHORS
Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: sds@sanger.ac.uk
On or before Oct 29, 2002 this sequence version replaced
gi:20520791, gi:20520768, gi:20520769, gi:20520793, gi:20520794,
gi:20520822, gi:20520770, gi:20520690, gi:20520671, gi:20520801,
gi:20520696, gi:20520807, gi:20520866, gi:20520790.
JOURNAL
FEATURES
Location/Qualifiers
source
1. 321250
/organism="Streptomyces coelicolor A3(2)"
/mol_type="genomic DNA"
/db_xref="taxon:100226"
complement(200..2215)
/gene="SC01964"
/note="synonym: SCC54.24c"
complement(200..2215)
/gene="SC01964"
/note="SCC54.24c, possible export associated protein, len:
671aa; prediction based on the predicted function of
```

surrounding genes and the short region (67-162aa) of similarity to TR:Q52357 (EMBL:U38624) terD, associated with phage inhibition, colicin resistance and tellurite resistance from *Serratia marcescens* Inch12 plasmid R478
blast scores: E=2.1e-12, Identities = 40/98 (40%), Positives = 55/98 (56%) which may represent an active site."

/codon_start=1
/transl_table=11
/product="putative export associated protein"
/protein_id="CAB38154.1"
/db_xref="GI:4468700"
/db_xref="GOA:Q9Z502"
/db_xref="SPTREMBL:Q9Z502"
/translation="MTAELTRGQNHPLPRARLEIRVSAGTPVVAATLGDENGTHGV
ERVVHCPAPLPLGLVEGRQAADHLDLDAVDPAVHRVSVLLALPAGGQGFARFGA
VAAPFVATLDLGDDEVASYITGLAESAVVALELYRROGAWKVRVAGQVAGGLAEL
LADQKLSOQLAATHIENAVSGLARISIPAPPAAPAHDPHGAMPAGGPGVPVPAS
FYDPKGSAGPGQTPGAPRPGGQPTPSGGEFATAGQSPAPAGGPDYSHRRPOS
AAPPFPPAPASPEGPARPVAGDAGTGWMEERLYNQVMGFEDLARTTAAYKSAVD
FADSRMEKLDQVLSDRSIGGQGDAAEAAARHGQVLSQAREVLDRLDVAQKVAEA
EVEVPALPAAPARNDPNVHAYRVPMEIPMALRLGLHLPEADRIPIMLIRLPIERG
LWDSGSASLDGSGFADSHERRIGLTAVSHAARLLAVYPAGFTVHVDPAGSGAQ
ALAPLAGSVLAAAPAGAGAADVLARLTCORVLDVOMALEGGAPDALPGLDPSQQL
LIVNDFPHGDRAVNRVYLADGEPAGVUHLMMVADRESAGFGPLDPLWRSLRL
TPVADDDHLPDWPWGHWMTYPSLVPFGSQVLQQLAQVAAARRSWDR"

gene
/length=2340. .2918
/size="SC01965"
/size="synonym: SCC54.25c"
/complement(2340. .2918)
/size="SC01965"
/size="synonym: SCC54.25c, possible export associated protein, len: 192aa; similar to many associated with plasmid borne resistance export systems egs. TR:Q52353 (EMBL:U38824) from the Inch12 plasmid R478 of *Serratia marcescens* (193 aa) fasta scores: opt: 497, z-score: 593.2, E(): 9.8e-26, (42.2% identity in 192 aa overlap) and SW:TERD_ALCSP_terD, tellurium resistance protein from the pMER610 plasmid of *Alcaligenes* sp. (192 aa) fasta scores: opt: 357, z-score: 429.6, E(): 1.3e-16, (37.2% identity in 188 aa overlap)."

/codon_start=1
/transl_table=11
/product="putative export associated protein"
/protein_id="CAB38155.1"
/db_xref="GI:4468701"
/db_xref="GOA:Q9Z501"
/db_xref="SPTREMBL:Q9Z501"
/translation="MTVNMWTKGQALSLQKSDGSLTSVRMGLGMQAPRRGLFGSRTR
EIDLDASVLADQPDVDFRHLVSDGSRVHTGDLNVLGGVGGGDEAILVDLQR
VPVHIDQIVFTNSFTGTFQEVQNAFCRLVDETNQGLARYTLAGGGAFTAQIMAKV
HRAGQGSMTAIGTPANGRTFDLMPAILPLV"

gene
/length=3061. .5199
/size="SC01966"
/size="synonyms: SCC39.01c, SCC54.26c, uvrB"
/complement(3061. .5199)
/size="SC01966"
/size="synonyms: SCC54.26c, partial CDS, probable uvrB, ABC excision nuclease subunit B, len: >113aa; similar to many eg.
SW:UVRB_MICLU_uvrB, ABC excision nuclease subunit B from *Micrococcus luteus* (709 aa) fasta scores: opt: 248, z-score: 280.5, E(): 2.6e-08, (45.9% identity in 111 aa overlap).
SC39.01c, uvrB, excinuclease ABC subunit B (fragment), len: >635 aa; highly similar to SW:UVRB_BACSU (EMBL:AF017113) *Bacillus subtilis* excinuclease ABC subunit B (DinA protein) UvrB or DinA or Uvr aa; fasta scores: opt: 2591, z-score: 2687.6 bits; 507.5 E(): 4.3e-142; 63.66% identity in 600 aa overlap. Contains Pfam matches to entries PF00270 DEAD, DEAD/DEAH box helicase and PF00271 helicase_C, Helicase conserved C-terminal domain and match to Prosite entry PS00017 ATP/GTP-binding site motif A (P-loop). Also contains a possible coiled-coil region at approx. residues 273. .230"

/codon_start=1
/transl_table=11
/product="ABC excision nuclease subunit B"
/protein_id="CAD55287.1"
/db_xref="GI:24419016"
/db_xref="GOA:Q8CK11"
/db_xref="SPTREMBL:Q8CK11"
/translation="MRPVQIERTVAFVFFVVPQSPGDOPTAIAELARVQAGEKDV
VLIGATGKSATTAEMIEKQPTLVNPNKTLAAQANEFRELLPNNAVEYFVSYY
DYQPEAYVPQSDTYTEKSSINNEEVRHSATNSLLTRRDVIVASVSCYIGLGP
QYVDVRMPLRVGEEHDELLRFRVDIOTRNDMAFARGTFRVGTDTIEFFPYEEL
AVRTEMFGDRIEALTLHPVTGETISDQOLYFPASHYVAGPERLRVANDIKPTEL
ERLTELKQKLLAEQALRMRTYVDIEMLQIGSCSVENYSHMFDCSGSPSNTLL
DYFPDILLIDESHVIVPQIGAMEGDASKRLVHGRPLPSALDNRELKWEFOE
RIGOTVYLSATPAGYELSDSGAVEIIRPTGLVDVPPVVKATPEGQIDDLVHEIRRT
EKORVILVTLTKMAEDLTDYVELGIQVRYLHSDVTLRRLVRELRLAGEYDVLV
GINLLRGLDPELSVLAIDADKEGFLRGSTLSIOTIGRAARNVSGVHMVADKITP
AMEKAIDNRRREKQVAFNKAQGVDPQLRKINDIQAOTAREDDVTEQLLGSYRQ
TKGKGAKAPVALGGOKTGAKAARGKETAVTDRAPAELAQIEDLITTRMRAAAA
DLQFEIARLRDEVSEMKELRQWREAGLA"

misc_feature
/length=3541. .3801
/size="SC01966"
/size="pfam match to entry PF00271 helicase_C, Helicase conserved C-terminal domain, score 75.00, E-value 1.5e-18"
/complement(4753. .5166)
/size="SC01966"
/size="pfam match to entry PF00270 DEAD, DEAD/DEAH box helicase, score 27.00, E-value 1.2e-07"
/complement(5035. .5058)
/size="SC01966"
/size="PS00017 ATP/GTP-binding site motif A (P-loop)"
/complement(5239. .6117)
/size="SC01967"
/size="synonym: SC39.02c"
/complement(5239. .6117)
/size="SC01967"
/size="SC39.02c, possible integral membrane protein, len: 292 aa; similar to TR:Q9KZ14 (EMBL:AL353863) Streptomyces coelicolor putative integral membrane protein SCG8A.13c, 244 aa; fasta scores: opt: 703, z-score: 726.3 bits; 142.1 E(): 7.6e-33; 48.430% identity in 223 aa overlap. Contains possible hydrophobic membrane spanning regions"

/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAC42741.1"
/db_xref="GI:14571767"
/db_xref="SPTREMBL:Q93J04"
/translation="WHSTVQGFYGLVTELVAYLMACLGALGLCTARAVLVRSWR
PCWLALGSAALSGIWMFHMFGVMMGFIEHTPIRYDMLTASLAVAVMVGVIPIV
GYRGARGTALFTGGTITGLGIAASHYLGMAHGLDGGQTYNTFTTAVSVVLAAMVATA
ALWAAQVQVGFELVSGAALINGLAVTGMHTYMAAEVHLVLSGTAEPSVGGSPAEALLAP
MLIGLAPLLLAGVVVVMFPLVMVGRPARVPAERKPGIPATHTVPTIRHPAHRPRR
TRHPLGHRRSRTPQNR"

RBS
6399. .6403
6409. .7278
/size="SC01968"
/size="synonym: SC39.03"
6409. .7278
/size="SC01968"
/size="SC39.03, probable secreted hydrolase, len: 289 aa; similar to SW:GLFQ_BACSU (EMBL:Z26522) *Bacillus subtilis* glycerophosphoryl diester phosphodiesterase (EC 3.1.4.46) GlpD, 293 aa; fasta scores: opt: 495, z-score: 576.9 bits; 114.7 E(): 1.6e-24; 34.114% identity in 299 aa overlap. Contains possible cleavable N-terminal region signal peptide sequence"

/codon_start=1
/transl_table=11
/product="putative secreted hydrolase"
/protein_id="CAC42742.1"
/db_xref="GI:14571768"
/db_xref="GOA:Q93J03"

```
/db xref="SPTREMBL:Q93J33"
/translations="MHVRAVAVTTTLLGLVALTAPLSHARADBGDDPTVVAHRGAS
GYAPENTLAAVDAAELGIRVENDVQRTDGLVILHDSLARTDVEVFDRSPW
KVDFPTAAELARLDAGSWFSFAGARVPTLEQYVRVHKNQKLLLEKSPGIPGI
EQOTLKVLNENEGWLDHRHVRAGRLVQSFDDASIRTVHDLKPAVKTFGLGTPAVSELP
Query Match      20.8%; Score 54.4; DB 1; Length 321250;
Best Local Similarity 52.7%; Pred. No. 6;
Matches 118; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY      4  CAGAGTGGAGTTTCGAGAGCTGTCGACACAGACGCCGCGCGCTGTCGAGAGGGCT 63
Db      204911  CGAGGCGCGCACGACGTCGCCGCGGTGGACGCTGCGACGATCATCGCGCGCGCAT 204970

QY      64  GAGCCTGCGGAGCTGTCTCTGCGGAGAGCGTGCCTTCATCAGACCATCCGGTCTGT 123
Db      204971  GGCCTTGGCGAGCGCTGTCTGCGCGCTCCGAGCTCCTCAGGACGCCAGCAACGGC 205030

QY      124  CGCGCCAGTCTGTCCTCGCCACCGCGGAGCGCCGATGCGCCTCAAGGCGAGCGCTGCC 183
Db      205031  GCTGCGGATGCTGCTGCTGGGCGCCGCTTGGCGCGCGCGCTCAGCGCGGTGCC 205090

QY      184  CGCGCGCTGCCGAGAGCTGCGCTTCGAGGCGGAGGTGGAGT 227
Db      205091  GGTACCGGTGCGCTCCCGCGGACGCGCGGTGCGCGGCGCT 205134
```

RESULT 13

```
AC134262
LOCUS      Rattus norvegicus clone CH230-1K3, *** SEQUENCING IN PROGRESS ***,
DEFINITION 39 unordered pieces.
AC134262
VERSION    AC134262.1 GI:23307916
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     Rattus norvegicus
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 67702)
AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Delaney,K.R., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B.,
            Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,X., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Mitchell,J., Newton,N., Nguyen,A., Nguyen,N.,
            Oragunye,N., Nickerson,E., Nwokkwo,S., Ogih,M., Okwuonu,G.,
            Oquaye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
```

```
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
```

Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Umani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 67702)
Worley, K.C.
Direct Submission
Submitted (25-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: TUG3
Center clone name: CH230-1K3
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 28204 bases at least Q40
Consensus quality: 31728 bases at least Q30
Consensus quality: 33612 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 39 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1374: contig of 1374 bp in length
* 1375 1474: gap of unknown length
* 1475 2540: contig of 1066 bp in length
* 2541 2640: gap of unknown length
* 2641 3702: contig of 1062 bp in length
* 3703 3802: gap of unknown length
* 3803 4826: contig of 1024 bp in length
* 4827 4926: gap of unknown length
* 4927 6309: contig of 1383 bp in length
* 6310 6409: gap of unknown length
* 6410 7910: contig of 1501 bp in length
* 7911 8011: gap of unknown length
* 8011 9211: contig of 1200 bp in length
* 9211 9310: gap of unknown length
* 9311 10837: contig of 1527 bp in length
* 10837 10937: gap of unknown length
* 10938 12361: contig of 1424 bp in length
* 12362 12461: gap of unknown length
* 12462 13981: contig of 1520 bp in length
* 13982 14081: gap of unknown length
* 14082 15311: contig of 1230 bp in length
* 15312 16475: contig of 1064 bp in length
* 16476 16575: gap of unknown length
* 16576 17971: contig of 1396 bp in length
* 17972 18071: gap of unknown length
* 18072 19486: contig of 1415 bp in length
* 19487 19586: gap of unknown length
* 19587 20912: contig of 1326 bp in length
* 20913 22715: contig of 1703 bp in length
* 22716

* 22716	22815:	gap of unknown length	
* 22816	24016:	contig of 1201 bp in length	
* 24017	24116:	gap of unknown length	
* 24117	25458:	contig of 1342 bp in length	
* 25458	25558:	gap of unknown length	
* 25559	27711:	contig of 2153 bp in length	
* 27712	27811:	gap of unknown length	
* 27812	29329:	contig of 1518 bp in length	
* 29330	29429:	gap of unknown length	
* 29430	31208:	contig of 1779 bp in length	
* 31209	31308:	gap of unknown length	
* 31309	32470:	contig of 1162 bp in length	
* 32471	32570:	gap of unknown length	
* 32571	33957:	contig of 1387 bp in length	
* 33958	34057:	gap of unknown length	
* 34058	35757:	contig of 1700 bp in length	
* 35758	35857:	gap of unknown length	
* 35858	37951:	contig of 2094 bp in length	
* 37952	38051:	gap of unknown length	
* 38052	39967:	contig of 1916 bp in length	
* 39968	40067:	gap of unknown length	
* 40068	41995:	contig of 1928 bp in length	
* 41996	42095:	gap of unknown length	
* 42096	43347:	contig of 1252 bp in length	
* 43348	43447:	gap of unknown length	
* 43448	45228:	contig of 1781 bp in length	
* 45229	45328:	gap of unknown length	
* 45329	47658:	contig of 2330 bp in length	
* 47659	47758:	gap of unknown length	
* 47759	49506:	contig of 1748 bp in length	
* 49507	49606:	gap of unknown length	
* 49607	51308:	contig of 1702 bp in length	
* 51309	51408:	gap of unknown length	
* 51409	53679:	contig of 2271 bp in length	
* 53680	53779:	gap of unknown length	
* 53780	55623:	contig of 1844 bp in length	
* 55624	55723:	gap of unknown length	
* 55724	57057:	contig of 1334 bp in length	
* 57058	57157:	gap of unknown length	
* 57158	59462:	contig of 2305 bp in length	
* 59463	59562:	gap of unknown length	
* 59563	61452:	contig of 1890 bp in length	
* 61453	61552:	gap of unknown length	
* 61553	64604:	contig of 3052 bp in length	
* 64605	64704:	gap of unknown length	
* 64705	67702:	contig of 2998 bp in length.	
FEATURES			Location/Qualifiers
source			1..67702
			/organism="Rattus norvegicus"
			/mol_type="genomic DNA"
			/db_xref="taxon:10116"
			/clone="CH230-1K3"
ORIGIN			
Query Match			19.8%; Score 51.8; DB 2; Length 67702;
Best Local Similarity			52.0%; Pred. No. 22;
Matches 116; Conservative			0; Mismatches 107; Indels 0; Gaps 0;
QY	4	CAAGTGGAGTTTCGAGGAGCTGTCGAGACCGGCGGCGCTGCTGGAGGGCT	63
Db	33568	CATGGCGAGCTGTACGAGCGGAGGTGCGGAGATGCGGGGCGTGTGGCTGGC	33627
QY	64	GAGCCTCGGGGACGTGTTCTCGGCGAGACGGTGCCCTTCATCAAGACCATCGGCTGT	123
Db	33628	GGCGGCGCGGCGAGCTGCGCTGAGCAGGAGCACTTGCTGGAGGACATCGGCACGT	33687
QY	124	GGGCGAGTGTGCTGCTCGCCACCGGAGCGCCGATGCGCCCTGAAGGAGCGCGTGGC	183
Db	33688	CGGCCAGCGGCTGGAGGAGGAGCGCGGAGCGCGGAGGAGCGGCGGCGCGCGC	33747
QY	184	CGCGCGCTCCCGGAGAGTGSCCTTCGAGGCGGAGGTGGAG	226
Db	33748	CCTAGCCCGCTTCGCGCAGAGCGGAAGCGCGCGCGTGGAG	33790

RESULT 14

AC118991

LOCUS

DEFINITION

AC118991

AC118991.4

HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.

Rattus norvegicus

SOURCE

Rattus norvegicus

ORGANISM

Rattus

1 (bases 1 to 219992)

Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,

Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,

Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,

Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,

Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,

Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,

Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,

Chacko,J., Chavez,D., Chen,G., Chen,Y., Chen,Z., Chu,J.,

Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,

Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,

Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,

Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,

Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,

Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,

Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,

Gebregregis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,

Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,

Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,

Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,

Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,

Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jollivet,A.,

Karpatis,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,

Kowis,C., Kraft,C.L., Lebow,H., Levay,J., Lewis,L., Li,Z., Liu,J.,

Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,

Lorensuhewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,

Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,

Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,

Mawney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,

Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,

Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,

Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,

Nwackelmech,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,

Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankuch,C.,

Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu.L.-L.,

Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,

Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,

Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,

Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,

Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,

Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,

Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabok,P., Taylor,C.,

Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,

Valas,R., Vera,S., Villalana,D., Waldron,L., Walker,B., Wang,J.,

Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,

Williams,G., Willson,R., Wleczek,R., Wooden,H., Worley,K.,

Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,

Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von

Niederhauser,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,

Weinstock,G. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 219992)

Worley,K.C.

Direct Submission

Submitted (24-APR-2002)

Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 219992)
Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 6, 2002 this sequence version replaced gi:23682583.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GTCW
Center clone name: CH230-142114
----- Summary Statistics

Assembly program: Phrap; version 0.990329
Consensus quality: 270777 bases at least Q40
Consensus quality: 209387 bases at least Q30
Consensus quality: 210977 bases at least Q20
Estimated insert size: 211380; sum-of-contigs estimation
Quality coverage: 9x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 218703: contig of 218703 bp in length
* 218704 218803: gap of unknown length
* 218804 219992: contig of 1189 bp in length.

Location/Qualifiers

1. .219992
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-142114"

FEATURES

source

misc_feature
1. .1360
/note="wgs end extension
clone end:Sp6"
2228..3263
/note="wgs end extension
clone end:Sp6"
misc_feature
10120..10615
/note="clone boundary
clone end:Sp6
site:
end sequence:RWAXN55TVB"

misc_feature

201059..217135
/note="clone boundary
clone end:17
site:
end sequence:RWAXN55TVB"

ORIGIN

Query Match

19.8%; Score 51.8; DB 2; Length 219992;

Best Local Similarity 52.0%; Pred. No. 17;
Matches 116; Conservative 0; Mismatches 107; Indels 0; Gaps 0;
QY 4 CAAGGTGGAGTTCGAGGAGTGTCTGACACGCGCGCGCGCTGCTGAGGGGCT 63
Db 82343 CATGGGCGAGCTGTACGAGCGGAGGTGCGGAGATGCGGGGCGCTGCTGCGCCTGGG 82402
QY 64 GAGCTTGGGGGAGCGTGTCTTCTGGGCGGAGCGGTGCTTCTATCAAGACCATCCGGCTCGT 123
Db 82403 GCGCGGCGCGGGGCGAGCTGCGGCTTGGAGCAGGAGCACTTGTGTGAGGACATCGGCACGT 82462
QY 124 GCGGCCAGTCTGCTCCCTCGCCACCGCGGAGCCGAGTGGCCCTCAAGAGGAGGCGCTGCC 183
Db 82463 CCGCCAGCGCTGACGAGGAGGCGCGGAGCGGAGGAGCGGAGGCGGCGCGCGC 82522
QY 184 CGCGCGCTGCGCGGAGGAGTGGCTTCTGAGGCGGAGGTGGAG 226
Db 82523 CCTAGCCGCTTCTGCGCAGGAGGCGGAAGCGCGCGGTGGAG 82565

RESULT 15

AC095429

LOCUS

DEFINITION

AC095429

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Altschrocks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, N., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenschew, L., Loulseg, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankweli, O., Okwuonu, G., Olarnpungoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pu, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D.,

